HZ BOOKS
华章教育

国外电子与电气工程技术丛书

WILEY

概率、随机变量和随机过程在信号处理中的应用

[美] 约翰 J. 申克（John J. Shynk）著
谢晓霞 安成锦 许可 译

Probability, Random Variables, and Random Processes
Theory and Signal Processing Applications

PROBABILITY, RANDOM VARIABLES, AND RANDOM PROCESSES
Theory and Signal Processing Applications
JOHN J. SHYNK
WILEY



机械工业出版社
China Machine Press

本书是关于概率理论的综合指南，其中总结了大量关键结论，扩展了各种可应用到信号处理领域的统计技术。本书是一本适用于工程师学习概率原理的书，它给出了一个比本科生通常使用的相关书籍更严格的数学框架。它可用于对概率和随机变量较熟悉，但不一定涉猎过随机过程和随机信号作用下系统的一年级研究生，也可用于具有很强数学背景的高年级本科生，以及信号处理领域的工程师。

本书特色

- 附录不仅包括积分、重要的不等式和恒等式、频域变换以及线性代数，还包括对33个随机变量及其特性的总结，将这些内容包括进来使得本书相对独立。
- 本书图文并茂，结合六百多张插图和Matlab图来对重要的知识点进行详细说明。
- 详细讨论充分统计量以及它们与参数估计的联系。包括经典的贝叶斯估计和几个最优准则：均方误差、平均绝对误差、最大似然法、矩方法、最小二乘法。

作者简介

约翰 J. 申克（John J. Shynk）加州大学圣巴巴拉分校电气和计算机工程学院的教授。他曾是贝尔实验室的一名技术人员，在波士顿大学与斯坦福大学获得系统工程、电子工程和统计学位。他的研究方向包括：自适应信号处理、通信和生物医学工程领域的信号处理等。





WILEY
www.wiley.com

投稿热线：(010) 88379604
客服热线：(010) 88378991 88361066
购书热线：(010) 68326294 88379649 68995259

华章网站：www.hzbook.com
网上购书：www.china-pub.com
数字阅读：www.hzmedia.com.cn

上架指导：信号处理

ISBN 978-7-111-51965-2



定价：129.00元

封面设计·锡彬

国外电子与电气工程技术丛书

概率、随机变量和随机过程在信号处理中的应用

[美] 约翰 J. 申克（John J. Shynk）著
谢晓霞 安成锦 许可 译

Probability, Random Variables, and Random Processes
Theory and Signal Processing Applications

PROBABILITY, RANDOM VARIABLES, AND RANDOM PROCESSES
Theory and Signal Processing Applications
JOHN J. SHYNK
WILEY



机械工业出版社
China Machine Press

图书在版编目（CIP）数据

概率、随机变量和随机过程在信号处理中的应用 /（美）申克（Shynk, J. J.）著；谢晓霞，安成锦，许可译 . —北京：机械工业出版社，2016.2
（国外电子与电气工程技术丛书）
书名原文：Probability, Random Variables, and Random Processes：Theory and Signal Processing Applications

ISBN 978-7-111-51965-2

I. 概… II. ①申… ②谢… ③安… ④许… III. ①概率论 – 应用 – 信号处理 ②随机变量 – 应用 – 信号处理 ③随机过程 – 应用 – 信号处理 IV. TN911.7

中国版本图书馆 CIP 数据核字（2016）第 027936 号

本书版权登记号：图字：01-2013-4261

本书包括概率与随机变量、随机过程、随机变量与随机过程在信号处理和通信中的应用三部分。通过大量内容来支撑和扩展本书重点，用 600 多个实例和 MATLAB 来增强和阐明随机量的特征和特性，翔实的统计数据将经典的贝叶斯估计和一些最优性准则用于参数估计技术。最后着重阐述了随机过程与系统在通信系统和信息理论、最优滤波（维纳滤波和卡尔曼滤波）、自适应滤波（FIR 和 IIR）、天线波束形成、信道均衡和测向中的实际应用。

本书适合作为有一定概率与随机过程理论基础的电子信息和通信专业的研究生教材，也可用于具有很强数学背景的高年级本科生。

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）
责任编辑：谢晓芳　　责任校对：殷　虹
印　　刷：北京诚信伟业印刷有限公司　　版　　次：2016 年 4 月第 1 版第 1 次印刷
开　　本：185mm × 260mm　1/16　　印　　张：34
书　　号：ISBN 978-7-111-51965-2　　定　　价：129.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzjsj@hzbook.com

出版者的话

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，信息学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的信息产业发展迅猛，对专业人才的需求日益迫切。这对我国教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其信息科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀教材将对我国教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到"出版要为教育服务"。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与Pearson、McGraw-Hill、Elsevier、John Wiley & Sons、CRC、Springer等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出Alan V. Oppenheim Thomas L. Floyd、Charles K. Alexander、Behzad Razavi、John G. Proakis、Stephen Brown、Allan R. Hambley、Albert Malvino、Peter Wilson、H. Vincent Poor、Hassan K. Khalil、Gene F. Franklin、Rex Miller等大师名家的经典教材，以"国外电子与电气技术丛书"和"国外工业控制与智能制造丛书"为系列出版，供读者学习、研究及珍藏。这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版"经典原版书库"作为姊妹篇也越来越多被实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着电气与电子信息学科建设的不断完善和教材改革的逐渐深化，教育界对国外电气与电子信息教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下：

华章网站：www.hzbook.com

电子邮件：hzjsj@hzbook.com

联系电话：(010)88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



华章科技图书出版中心

译者序

John J. Shynk 教授，1986 年从美国斯坦福大学获得博士学位，现任美国加州大学圣巴巴拉分校电气和计算机工程系教授，致力于自适应信号处理技术在通信和生物工程方面的应用研究。作者以多年的教学和科研经验为基础撰写了本书，通过对测度论、域以及对随机过程的相关性和它的用途的深入探讨提供了对概率、随机变量和随机过程的更严格的数学框架。本书第 1 章是全书的概述以及相关背景知识的介绍，使读者在一开始就对本书将要介绍的内容有基本了解。之后全书分为三部分，分别讨论了概率、随机变量和数学期望；随机过程、系统和参数估计以及这些概率理论在信号处理和通信中的应用。书中给出了大量的例题和图表，对各种方法进行总结，对不同参数的结果进行对比分析，加深读者对各种定理、定律的理解。其他基础教材中常被忽略或没有充分说明的许多问题在本书中也有较详细的解释。书中还提供了研究数字通信、信息论、自适应滤波、线性和非线性估计和检测等课题的必要的背景知识，并在附录中提供了相关的数学和信号与系统的背景材料，使本书相对独立。

本书针对工程领域的一年级和二年级研究生，也可用于在概率与随机过程方面具有相关数学背景的高年级本科生。作者在前言中给出了针对不同课时的详尽的教学建议供参考。

参加本书翻译工作的有谢晓霞(前言、符号说明、第 6 章、第 7 章、第 9 章，附录 A～D，术语表)、安成锦(第 4 章、第 5 章、第 8 章、附录 E～G)、许可(第 1～3 章)。最后由谢晓霞对全书的译文进行审校。

由于译者水平有限，书中难免有不当之处，敬请读者批评指正。

前　言

本书旨在给出概率、随机变量和随机过程的一个数学框架，它比大多数本科院校普遍使用的概率和统计教材更严格。本书专门为一年级的研究生设计，如果教师忽略其中一些更关注理论的部分，则本书也可用于具有很强数学背景的高年级本科生。本书有以下几个特点：

- 给出大量总结各种技术和显示例题结果的详细的图和表，包括利用 MATLAB 产生的 600 多个插图，旨在使本书更有说服力。
- 包括对基础教材中没有涉及的许多问题进行说明的例子。每章还包含家庭作业，并提供答案给教师。
- 附录中提供的相关的数学和信号与系统的背景材料，使这本书相对独立。附录 A 给出一些参数的单变量分布的总结。
- 本书第三部分介绍了信号处理和通信应用，这些都是学生在以后的工程类课程中可能遇到的问题。这些介绍性的材料基于美国加州大学圣巴巴拉分校开设的几门统计课程。

第 1 章包括全书的概述以及线性系统和频域变换的介绍。全书后面分为三部分，每部分包括四章：

- 第一部分(第 2～5 章)：概率论；随机变量；多维随机变量；期望和矩。
- 第二部分(第 6～9 章)：随机过程；随机收敛、微积分和分解；系统、噪声和谱估计；充分统计量和参数估计。
- 第三部分(第 10～13 章)：通信系统与信息论；最优滤波；自适应滤波；均衡、波束赋形和测向。这四个章节位于 www.wiley.com/go/randomprocesses。[⊖]

第 2 章介绍了基本的概率，重点介绍的是离散试验。介绍了样本空间、事件和域，提供了抽象概率空间$\{\Omega, \mathcal{F}, P\}$的框架，它将用于后面介绍的随机变量和随机过程中。第 3 章定义了随机变量，其中包括许多著名的(和不那么知名的)连续和离散参数分布族的说明。第 4 章关注多维随机变量，研究了用于推导随机变量变换的分布的几种技术，还包括一些多元分布。第 5 章定义了随机变量的期望，以及随机变量函数的期望、矩和特征函数。还讨论了条件期望和它的几个重要性质。

通过将多维随机变量的特征扩展为由时间索引，第 6 章引入了随机过程。该章涵盖了随机过程的各种性质，如独立性和平稳性，描述了不同类型的随机过程，包括独立序列、马尔可夫链和鞅。深入分析和研究了大家熟知的随机过程，例如泊松和维纳过程。第 7 章

⊖ 关于本书教辅资源，用书教师可向约翰·威利出版公司北京代表处申请，电话：010-84187815，电子邮件：iwang@wiley.com。——编辑注

探讨了随机过程的其他特征，包括随机连续、导数、积分和微分方程。还描述了随机序列的随机收敛，以及大数定律和中心极限定理。第 8 章定义了随机过程的功率谱密度，它可用于描述经过系统处理(滤波)后的信号特性。该章重点讨论线性时不变系统(尽管有时会提及一些非线性处理)。该章还讨论了用参数和非参数技术进行谱估计的方法。第 9 章介绍了充分统计量，描述了几个用于估计随机变量的参数的重要方法。这些技术都是基于各种准则，包括均方误差、最大似然和最小二乘法。

本书的最后部分从第 10 章对数字通信的概述开始，其中包括对信息论的介绍。推导出了基于最大后验概率和最大似然准则的检测器。第 11 章分析了最优滤波技术，重点讨论了均方误差准则，推导出因果和非因果维纳滤波器，还介绍了采用格型滤波器进行线性预测和基于状态空间模型的信号的卡尔曼滤波器。第 12 章描述了自适应滤波算法和结构，从最陡下降法和牛顿法的讨论开始。还描述了研究自适应算法的收敛性及其稳态特性的随机方法。最后，第 13 章介绍了自适应波束的形成，其中用多个天线收集和分离同信道信号。还讨论了自适应均衡技术，它用来补偿传输信道中的信号失真。我们描述了理想的和基于训练的方法，以及不需要训练或导频信号的“盲”算法，还介绍了用于估计照射到天线的信号到达角的测向算法。

七个附录提供了书中所介绍主题的更多的背景材料，它们包括：

- 22 个连续和 11 个离散的随机变量的单变量分布的总结。
- 整本书用到的函数的连续性和具有特定符号的几个函数的描述。
- 离散和连续时间的频域变换，包括性质表和一些变换对。
- 黎曼积分的回顾，黎曼-斯蒂尔切斯和勒贝格积分的简要说明，以及有用的不定积分和定积分的总结。
- 恒等和无穷级数，主要用于离散型随机变量和随机序列。
- 期望值的边界和不等式的推导，如马尔可夫和切比雪夫不等式以及克拉美罗下界。
- 矩阵的几个性质，包括子空间、向量的分解和微分。

读者会发现在课程的学习过程中以及对微积分、信号与系统和线性代数课程的回顾过程中，附录中的内容很有用。

本书可用于整个学年，每季度覆盖前面提到的三部分之一：

- 学季制。秋季：第 1～5 章；冬季：第 6～9 章；春季：第 10～13 章。

对学期制，它可以分开学习，这样系统、估计和应用的内容就可以到第二学期再学习：

- 学期制。秋季：第 1～7 章；春季：第 8～13 章。

这本书也可在一季度，两季度或通过省略几个章节中的一些较高深的内容来在一学期学完。对于一季度的课程(10 周教学)，教师应该能够覆盖第 1～8 章中大部分内容，例如可以省略随机积分和谱估计。对于一学期的课程(15 周教学)，可以包括第 9 章充分统计量和参数估计的内容，省掉前面章节的一些内容。应用章节可供有兴趣的学生作为以后学习其他工程课程的预览。在前面章节学习随机过程、系统和噪声时使用这些内容作为系统或

信号处理的例子是比较合适的。

感谢 S. Chandrasekaran 审校附录，J. D. Gibson 提供的支持。感谢修过我开设的 ECE 235、ECE 240A 和 ECE 245 课程的学生，他们的问题促成了本书几个专题细致入微的理解。他们提供的宝贵的反馈意见促成了一些讨论和说明性的例题。我还要感谢电气和计算机工程学院的同事们，与他们的互动和他们多年的见解使我对统计信号处理和其应用范围之广有了更深的了解。最后，要感谢出版商 Wiley，感谢 George Teleck 对这个项目的支持，以及 Kari Capone、Dan Timek、Stephanie Loh 和 Shalini Sharma 在本书出版的最后阶段的协助。

符号说明

由于本书所涵盖的内容范围非常广，因此在此给出关于符号的简要概述。在许多信号处理的书中，用同样的符号来表示一个随机过程和该过程的现实，过程的现实是一个确定的波形。在本书中，使用概率和随机过程的书中的典型符号来替代：

- 大写字母 X 表示随机变量、随机过程 $X(t)$或随机序列 $X[k]$，其中 t 是连续时间，k 是离散时间。
- 小写字母 x 是 X 的一个输出，$x(t)$是 $X(t)$的一个现实(连续波)，$x[k]$为 $X[k]$的一个现实(数列)。

这些字母通常来自拉丁字母表的末尾。上述符号中只有一个例外，就是大写的 K、M 和 N 通常表示(非随机的)整数，如以下随机变量和：

$$\sum_{m=1}^{M} X_m, \quad \sum_{n=1}^{N} Y_n \tag{1}$$

如果在某个特定问题中$\{K, M, N\}$变成随机变量，例如在随机求和的问题中，将专门提及它。

- 加粗的大写字母表示随机向量 $\boldsymbol{X}$、随机向量过程 $\boldsymbol{X}(t)$或随机向量序列$\boldsymbol{X}[k]$。
- 加粗的小写字母 $\boldsymbol{x}$ 是 $\boldsymbol{X}$ 的一个向量输出，$\boldsymbol{x}(t)$是 $\boldsymbol{X}(t)$的一个向量的现实(波形向量)，而 $\boldsymbol{x}[k]$为$\boldsymbol{X}[k]$的一个向量现实(向量序列)。

本书中所有的向量都是*列向量*。行向量通过转置 $\boldsymbol{x}^{\mathrm{T}}$，或复共轭的转置 $\boldsymbol{x}^{\mathrm{H}}$ 得到。$\boldsymbol{x}^{*}$ 的上标只表示复共轭，不包括转置。

- 加粗的大写字母 $\boldsymbol{A}$ 也用于非随机矩阵，加粗的小写字母 $\boldsymbol{a}$ 表示一个非随机向量。通常，这些字母来自拉丁字母表的开始。

读者应该能够从讨论的上下文中确定 $\boldsymbol{X}$ 是否是一个随机向量或一个非随机矩阵。两个重要的例子是自相关矩阵 $\boldsymbol{R}_{XX}$ 和自协方差矩阵 $\boldsymbol{C}_{XX}$，它们是随机向量 $\boldsymbol{X}$ 的非随机量。

- 英文花体$\mathcal{E}$用于表示期望。例如，自相关矩阵为 $\boldsymbol{R}_{XX} \triangleq \mathcal{E}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$。

虽然，在许多关于概率的书中 E 用来表示期望，但我们用$\mathcal{E}$，因为在某些章节中必须用 E 表示一个误差随机变量(E 也可以用来表示在样本空间 Ω 中的一个事件)。

为了使整本书的许多式子简单明了，分母中有时省略了小括号。例如，将表达式写为 $1/2\pi\mathrm{j}$，它应当被理解为 $2\pi\mathrm{j}$ 的所有项都在分母中，而不必使用括号 $1/(2\pi\mathrm{j})$。另一个例子是高斯概率密度函数(pdf)：

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/2\sigma^2) \tag{2}$$

它应当清楚地表明了 $2\sigma^2$ 是在指数的分母中(尽管这不是大多数计算机编程语言的操作顺序)。

在本书末尾的术语表中，给出了上述符号的总结，以及整本书中使用的符号和缩写的列表：(ⅰ)一般符号，(ⅱ)希腊符号(通常用于随机变量参数)，(ⅲ)手写体符号(对于特殊的量和前面提到的期望$\mathcal{E}$)，(ⅳ)数学符号，(ⅴ)缩略词。

目　录

第二部分 随机过程、系统与参数估计

附录 附录内容介绍

第1章 内容概述与背景知识

1.1 引言

任何一个物理过程，不管它是自然的还是人造的，都可以看作是一个具有时变特性的信号，可以用随机过程对这个信号建模。“信号”是指通过物理过程表示的信息，而“随机”是指过程的未来结果在某种程度上是不可预测的。自然界随机信号的例子包括：照射到树上面的紫外辐射，围绕着恒星旋转的行星，通过一片开阔地的龙卷风等。合成的随机信号包括：从手机发射到基站的微波信号，从洛杉矶行驶到旧金山的汽车，以及投掷手给接球手的棒球等。这些例子都可以通过函数 $g\{x,y,z,t\}$ 进行建模，该函数描述了一个在三维空间 $\{x,y,z\}$ 随时间 t 变化的轨迹。很显然，这些“信号”具有不同的物理机质，不同程度的可预测性以及包含了不同的信息量。图1-1就描绘了一个龙卷风在 x 方向上随时间变化的可能的轨迹(称为现实)。可以看出，在一维空间坐标上，一个复杂自然过程的轨迹也类似于信号的概念。

需要注意的是，测到的龙卷风的轨迹是龙卷风在 x 方向上运动的一个可能现实，该轨迹是参照某个特定的经线定义的。也就是说在观测实际的轨迹之前，在某个区域内会有许多(甚至无穷多)个现实。换句话说，如果我们能够“重启”这个过程，允许龙卷风再来一次的话(重复这个“试验”)，在不同的大气条件下，我们将期待有一个不同的轨迹，而且这个轨迹与第一次观测到的轨迹只有少许不同。用 $X(t)$ 来表示一个随机过程，用 $x(t)$ 来表示这个随机过程的现实。随机过程在时间 t_0 的取值也是随机的，它是一个随机变量 $X(t_0)$，而在 t_0 时刻的某个特定取值是 $x(t_0)$。随机过程是以时间为索引的随机变量的集合。图1-2给出了一个现实 $x(t)$ 以及 $t=t_0$ 的一个取值 $x(t_0)$。

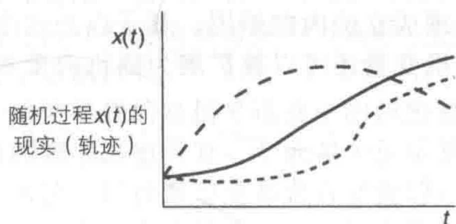


图1-1 随机过程在 x 方向上测到的随时间 t 变化的三条可能轨迹

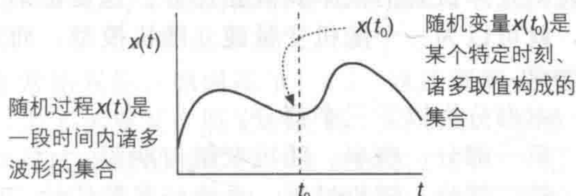


图1-2 随机过程 $X(t)$ 的现实 $x(t)$ 和随机变量 $X(t_0)$ 在 $t=t_0$ 时刻的取值 $x(t_0)$

很显然，我们无法用确定的方式预计龙卷风的轨迹，即使有大量的观测，如龙卷风的速度、地表温度、当天的时间、地理位置、地形等。这一点在下面的简单例子中会做进一步讨论：图1-3中描绘的均匀硬币的一次投掷。如果把投掷出去的硬币垂直立在桌面这种“极不可能”的情况排除，那么投掷硬币试验就只会有两种可能的结果：正面(H)和反面(T)。由于我们仅对硬币在一个特定时刻如何着陆(正面或反面)感兴趣，并不关心它的轨迹，因此单次投掷硬币试验可以用一个随机变量来建模，而不是一个随机过程。如果给定一些观测量，比如硬币着陆之前的速度、与桌面之间的入射角 ϕ、硬币的质量、大气条件(如温度、湿度



图1-3 通过物理方法来预测硬币将是正面(H)朝上还是反面(T)朝上

等)，再加上其他一些相关的物理参数，那么在硬币着陆之前就有可能预测出此次投掷试验的结果。要想可靠的预测硬币投掷试验的结果，似乎需要无穷多个这样精确的测量。但是，物体的速度和位置这两个参数是不可能同时准确测量到的，测量一个物体的位置就会改变它的速度(尽管这种改变非常微小)，反之亦然。事实上，我们不可能有关于随机事件的充分的物理观测来无差错地预测结果(除非只有一种可能的结果)。

通过前面的讨论可知，随机性可看作物理过程完整信息的缺失，以至于在随机事件发生之前我们无法精确预测它的结果或现实。因此，建立一个不直接依赖于特定过程的物理属性并且能够应用于信号和随机试验的随机模型是非常有用的。对于均匀硬币的投掷试验，凭直觉，随着试验次数的增加，H 和 T 出现的次数一样多。这种包含重复试验的观点称为概率模型的频率解释：在大量重复试验中，一个事件发生的频率决定了它的概率。如果硬币被投掷了 N 次，正面被观测到了 N_H 次，那么，对于 M 次硬币投掷试验，我们会期望正面朝上的次数大约为$(N_H/N)\cdot M$ 次。比值 N_H/N 是 H 出现的频率。同样的，T 出现的频率为 $N_T/N(N_T+N_H=N)$，我们也会期望在 M 次硬币投掷试验中 T 发生的次数大约为$(N_T/N)\cdot M$。

通过频率的方法来建立概率模型，与观测多次重复试验结果的可能性的方法是一致的。这个解释是基于三大公理建立概率模型的直觉基础。注意到 $0\leqslant N_H/N\leqslant 1$，即事件发生的频率总在区间 $[0,1]$ 之内。用 $P(E)$来表示事件 E 发生的概率。概率的一个公理是 $P(E)\geqslant 0$。这个下限的要求不仅来自基于频率的解释，而且也是数学的表示。另一个公理是，在一次试验中"事件发生"的概率为 1。综合上述两个公理可以得出：$0\leqslant P(E)\leqslant 1$，该结论同前面随机试验的频率解释是一致的。

第三个公理有些复杂，它涉及事件的组合，在一次试验中最重要的就是给感兴趣的任何事件赋予概率。例如，在硬币投掷试验中考虑事件发生的概率 $P(E=H$ 或 $T)$，很显然这个试验只会有两个输出(再次声明不考虑硬币立在地面的情形)。另外，这两个结果是互斥的，因为在同一次投掷中硬币不可能正面和反面同时朝上。我们发现硬币正面朝上或反面朝上的概率为$(N_T+N_H)/N$。概率简单相加是因为这两种试验结果是"非重叠的"(二者没有交集)。进一步分析可知，因为在这个试验中不可能有其他结果发生，因此我们肯定可以得出 $N_T+N_H=N(N$ 是投掷试验的总次数)，或者有$(N_T+N_H)/N=1$。最后一个结果表明互斥试验的概率具有相加性。这就是第三个公理成立的内在原因。基于以上三个公理，就可以为一个随机变量建立随机模型，而这个随机变量还可以被扩展为随机向量和随机过程。

本书分为以下三个部分：

第一部分　概率、随机变量与期望

第二部分　随机过程、系统与参数估计

第三部分　信号处理和通信中的应用

在本章引言的后面部分，我们将给出本书的一个概述。本书使用的符号参见"符号说明"，而且这些符号会在后续章节中给出定义。引言部分以投掷硬币试验为例，表明如何通过利用或改变随机信号的特性处理从简单的随机试验到复杂的随机过程和工程系统。

1.1.1 信号、信号处理和通信

为了实现某个目的，我们往往对随机信号以及信号处理的各种技术感兴趣。下面给出有关信号的一般性定义。

定义(信号)　**信号**是"包含"一个事件信息，并通过时间和空间传播的物理量。信号通过特定的媒介从一个物理地址传播到另外一个物理地址。

现代社会中存在着许多的信号，比如用于商业广播和有线电视的信号。其实，也可以认为宇宙中几乎所有东西都是信号。比如从超新星传播过来的光，毫无疑问可被视为一种信号，因为它提供了一个特殊事件的信息。电磁辐射可能是一种非常熟悉的信号模式，但

信号不必被限制为电磁辐射的形式。通过地球大气层的流星也可以被视作为一种信号，因为它可能提供地球相对于太阳的当前位置关系(比如一年中的哪个月)。一次地震产生的信号可提供地壳运动的相关信息，一次火山喷发也可以透露出地表深处的信息。

图 1-4 给出了信号模型三个基本元素的方框图，分别是：(ⅰ)信号源；(ⅱ)信号本身和信号传输的媒介(有可能是最重要的)；(ⅲ)感知信号的一个或多个传感器(用于观测或接收信号)。自然界的信号取决于物理世界的事件，比如前面提及的那些例子。其他例子还包括潮汐的涨落、云团的移动、森林中大树的倒下。最后一个例子经常用在这个著名的问题中，即，"如果森林中一棵大树倒下时没有人听到，那么这棵倒下的大树产生声音了吗?"倒下的大树以及由它对周围空气产生的扰动可以看作一种信号，这意味着另外一个物理事件可能刚刚发生(比如雷电击中了这个大树)。很显然，上述的这个问题仅仅关心信号模型中的传感器模块(如图 1-4 所示)。声音是耳膜感觉到的振动，如果没有人在现场的话那么声音就不会被听到。但是对此可能还会存在争议，因为周围的空气被倒下的大树扰动了，那么一个信号就会随之产生；另外，在大树倒下撞击地面的过程中还会产生额外的信号(振动)。

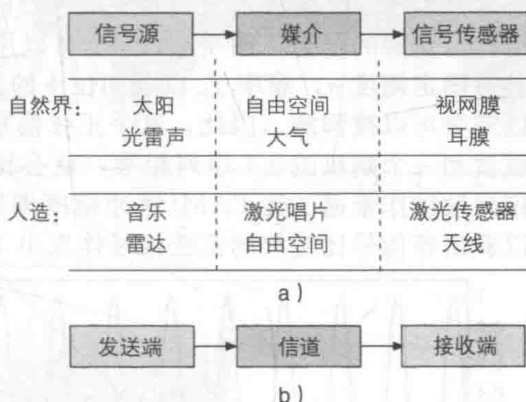


图 1-4　信号和通信。a) 信号传输的一般模型；b) 通信系统的典型模型

在此考虑那种飞越大洋的喷气式飞机，它的飞行可看作是在三维时空中的移动。大气湍流会导致它的飞行轨迹出现"随机"扰动，这使得它偏离"预定"轨迹。如果从上往下看它的飞行轨迹(不考虑 z 轴方向上的高度)，那么它在 xy 平面上的二维轨迹就可以看作时间 t 的函数。而且，如果只沿着 x 轴来观测它的轨迹，那么我们就会得到一个一维信号的现实(与前面提及的龙卷风的例子类似)。因此，即使是一个飞行的飞机也可以被看作一种形式的信号。如果这个"试验"能够被重复的话，那么后面的现实会与前面的有所不同，这是由于大气湍流和航向调整所致。因为传感器本身精度的限制以及实际情况中存在的噪声，这使得对信号的完美测量是不可能的，因此即使一个信号本身的随机扰动非常小，仍应该考虑这个影响。

这使得我们对信号模型中的传感器模块非常感兴趣，原因在于：(ⅰ)通过观测来建立信号的概率模型需要一个或多个传感器；(ⅱ)为了实现某个既定的目标，需要推导修改信号的方法。统计模型是非常有必要的，因为不可能通过足够多的测量来精确预测一个信号会如何演变。对于硬币投掷试验而言，尽管有可能通过非常多的物理测量来比较合理的预测试验结果(H 或者 T)，但该方法是不实用的。我们接受这个事实应该好一些，那就是：一个试验或信号是随机的，应该为试验的输出或现实建立一个概率模型，通过研究该模型来实现特定的应用。对于信号模型终端的传感器，它的随机性可视为对接收信号中不确定性的一种测量，因为我们不可能对生成和影响该信号的所有潜在的物理机理都非常了解。我们也不可能精确地知道信号在通过某些媒介的过程中受到的所有干扰。

图 1-4 也给出了一个跟信号有关的通信系统模型。该模型包括：(ⅰ)用于产生和发射信号的*发送端*；(ⅱ)*信道*，包括传输的媒介以及在信号被接收到之前的所有干扰；(ⅲ)*接收端*，包含一个或多个用于检测信号的传感器。该通信模型非常明确地认为某种形式的信息得到了传送和接收。在此考虑一种人为产生的信号，比如手机信号。很显然，打电话的人肯定知道传输信号中的所有信息。应该视随机性为接收端的一种特性，因为接电话的人并不知道打电话的人想说什么，同时由于信号在传输过程会有失真，接收端只能得到不完备的信息。该解释也同样适用于自然信号。例如假设存在一颗人造卫星，由于暴露在太阳

大气中所受到的太阳辐射强度会随着时间的变化而变化。该卫星在围绕地球旋转的过程中会把接收到的信号“视为”随机的(由不完备的信息所致)，通过对接收信号的处理有可能会获得该辐射所携带的信息(比如太阳上一个事件的发生)。

在第10章会介绍信息论的有关知识，一个信号携带信息的多少取决于不确定性的程度：一个具有“更多不确定性”的信号就会包含更多的信息。例如一个小行星由于受到彗星的猛烈撞击，它的轨道突然发生了剧烈的扰动，那么这个随机性就会提供比该小行星平稳运行时更多的信息。图1-5同样也可以用来说明这个情况(横坐标表示样本数目k)。一个具有固定幅度a，频率f_0以及相位ϕ的正弦信号是一个极端的例子，它是确定性信号，所以完全可以被预测。因此，这个正弦信号实际上并没有携带任何有用的信息。把一个正弦载波和一个随机的±1序列相乘，就会得到一个带通二元BPSK(相移键控)信号，该随机序列被称作带通二元PAM(脉冲幅度调制)信号。图1-5也给出了BPSK信号的波形，可以看出该信号比纯粹的正弦信号体现出了“更多的不确定性”。

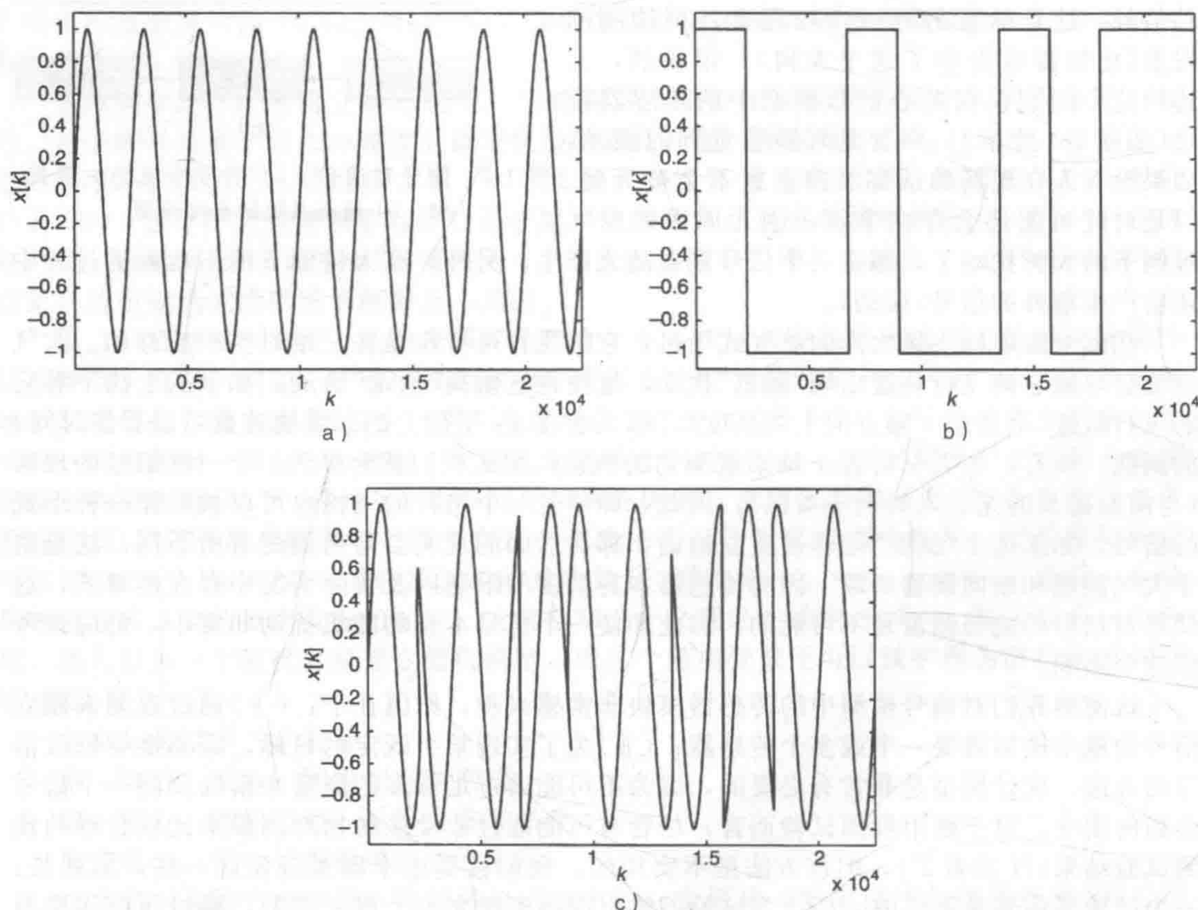


图1-5　人造信号的例子。a)具有固定幅度、相位和频率的正弦信号(非随机)；b)基带二元PAM信号(随机)；c)带通BPSK信号，由a和b中的信号相乘得到(随机)

图1-6中的语音信号体现了更多的随机性，因此它比前面的那些人造信号携带了更多的信息。图1-6还给出了风声的一个实现，这个信号结构更加简单而且跟前面的语音信号比起来具有更多的随机性。另外一个极端的例子是“白噪声”，这是完全不可预测的(后面将会对其进行定义)，该类型的信号是“最随机的”。我们也给出了一个视频截图的例子，这是一个二维的随机过程，该随机过程具有二维空间坐标$\{x,y\}$并随时间t的变化而变化。图1-7是对视频信号在一个特定时刻采集到的，因此它被建模为一个随机场。该图同时也给出了加入噪声的图片版本，这很直观地描绘了在传输过程中信道扰动带来的影响。信号处理技术通过去除一定程度的加性噪声，来改善信号的质量。

a）

b）

c）

图 1-6 自然界随机信号的例子。a）语音（Hu 和 Loizou，2007）；b）风声；c）高斯白噪声（通过 MATLAB 中的伪随机数发生器产生）



a） b）

图 1-7 随机场。a）来自一个视频的图像（单次截图）；b）噪声干扰的图像，每个像素点（图像的基本元素）都加上了零均值的高斯噪声

信息和信号的意义是有所区别的，信息是与前面提及的随机性相关的，而信号的意义更加复杂一些，这取决于接收终端如何解释信号。对于语音的例子，信号的意义取决于传输过程中使用的语言、接收端收听者的主要语言以及其他许多因素，比如说话者和收听者之间的关系、收听者以往的经验、说话者的口音等。本书并不关注信号的意义，尽管我们对蕴含其中的通过随机性来度量的信息感兴趣。我们也对各种各样的信号处理技术感兴趣，这些技术被用来提取和增强接收信号中的信息。

图 1-8 描述了两种我们将要讨论的基本信号处理方式，后面还会讨论各种各样的变化和应用。第一种是信号模型的参数估计，通过操作接收到的信号来设计出各种信号参数的估计器。最著名的估计器可能就是取样本均值：

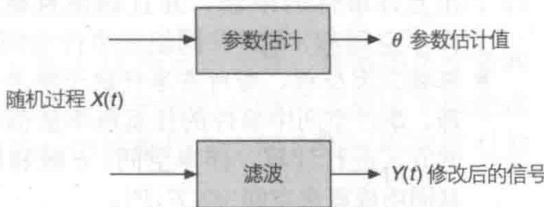


图 1-8 两种基本的信号处理方式

$$\overline{X}=\frac{1}{N}\sum_{k=1}^{N}X[k] \tag{1-1}$$

在这里通过对 $X(t)$(随机序列用 $X[k]$来表示)的 N 个样本取平均来估计信号的均值 μ_X。第二种信号处理方式是滤波，目的是为了变换信号，这样处理往往会“改善”它的性质。一般通过把信号和滤波器参数组$\{h[k]\}$相卷积得到：

$$Y[k]=\sum_{n=0}^{\infty}h[n]X[k-n] \tag{1-2}$$

一个信号处理的例子就是声音设计，将语音信号的幅度相乘，这样使得人说话的声音显得比真实记录的语音要响。将以上两种信号处理的模式进行组合同样是有可能的，我们可能需要在变换信号前对参数进行估计，那么就应事先用滤波器滤除背景噪声。

随机过程可被看作是一组以时间为标记的随机变量的集合，每个时刻都存在一个潜在分布的可能输出结果。下一步，我们给出随机模型的一个精确描述，并说明它是如何扩展为随机变量的。这使得我们可以采用微积分和离散数学的方法来*刻画*随机现象。通过引入时间变量的概念，这些方法被扩展到随机过程。如果在这些问题中考虑时间的话，我们往往会关心这些信号相邻时刻的*相关程度*。相关性可通过各种信号处理方法来进行研究，它是对随机信号将来输出预测的一种度量，甚至有可能是随机过程最重要的特性。

最后，我们讨论随机信号的空间维数。正如先前提及的那样，任何信号都可以看作一个随连续时间 $t(t\in\mathcal{R}=\{-\infty,\infty\})$ 或离散时间 $k(k\in\mathcal{Z}=\{\cdots,-1,0,1,\cdots\})$ 变化的三维变量(即 $\{x,y,z\}$ 空间坐标系)。但是，在很多情况下可能只需要从一维或二维空间上来考虑一个信号。对于先前那个行星围绕恒星旋转的例子，可能从椭圆平面观察这个过程更方便一些，这样就类似于观看一个(二维的)视频信号。对这个过程进行一次截图并取其对应的一帧信号，则与观看一次照片是基本等价的。全息摄影是一个三维的人造信号，从任何一个角度(平面视角)都是一个二维的视频，如果将其压缩到一个维度上，那么我们就只能观测到光强随时间的变化。绝大部分人造信号都是一维的，比如用于通信的信号。它们可能会在三维空间中传播，但它们会以各种现实的形式来表征，类似于图 1-5 中的那些波形例子。在本书中，我们主要关注的是一维的人造信号，但所讨论的技术完全可以扩展到任何类型的信号。

1.1.2 概率、随机变量和随机向量

接下来我们将归纳一些方法，利用这些方法可以推导出一个共同的*概率空间*。读者在学习本书前面章节以得出一个总体性的知识架构的时候，可能发现翻阅本节内容是有用的。

- *定义一个“试验”的输出值*。输出值是指一个随机试验的观测(可测)结果或对象。根据输出值的数量多少，可将输出值分为三大类：(i)有限的，(ii)无限可数的，(iii)不可数的。前两种对应离散的情况，第三种对应连续的情况。一个试验的所有输出值构成样本空间 Ω。
- *定义这个试验的事件*。事件 E 对应于单一的输出事件被称作基本事件。事件都是样本空间 Ω 的子集。
- *定义这些事件的代数集合，也称作 σ 域*。在许多问题中都会对事件和输出值进行集合运算，因此非常有必要定义一个 σ 域，通过这个域给出允许事件的限制，并且对事件赋予概率值。样本空间和 σ 域$\mathcal{F}$共同组成事件空间 $\{\Omega,\mathcal{F}\}$。
- *根据三大公理，给所有事件赋予概率值*。根据公理，事件空间中事件的任意概率值都可以用统一的方式进行计算。样本空间、σ 域和概率测度 P 共同组成概率空间 $\{\Omega,\mathcal{F},P\}$。

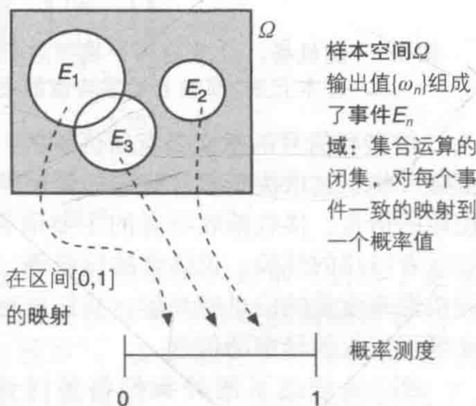


图 1-9 样本空间 Ω、事件$\{E_n\}$，以及概率值在区间 $[0,1]$ 映射的图形化描述。σ 域$\mathcal{F}$用来描述样本空间 Ω 中的事件，并对这些事件赋予概率值

图 1-9 形象地表示了这个框架结构。图中的矩

形表示样本空间 Ω，圆圈代表事件$\{E_n\}$，这个图就是所谓的维恩图。图中重合的那些部分是指具有共同输出值$\{\omega_n\}$的那些事件。由于定义了一个 σ 域$\mathcal{F}$，对于$\mathcal{F}$中所有事件的运算得到的仍然是$\mathcal{F}$中的一个事件。事件的运算包括$\cup$(或)，$\cap$(与)，^{c}(补)，$-$(非)，以及$\oplus$(异或)。根据 σ 域和三大公理，我们可以定义统一的概率测度 P，P 在闭区间 $[0,1]$ 中进行取值。

前面的这些内容适用于较小的样本空间。在实际中经常会遇到更为复杂的情况，此时就非常有必要把概率空间映射为实数直线$\mathcal{R}$上的随机变量。这样我们就可以对这些随机变量进行复杂的运算，比如把一个随机变量转换为另外一个随机变量。我们还可以定量的求出一个随机试验的各种特征，比如矩(均值、方差等)。常见的映射归纳如下：

- 将样本空间 Ω 的输出值映射为实线$\mathcal{R}$上的数值，得到随机变量 X。因为在工程问题中习惯用变量和数值，所以需要把原始样本空间 Ω 中的输出值映射为$\mathcal{R}$(或者整数$\mathcal{Z}$，复数$\mathcal{C}$)。原始的概率空间转移到了实线上，并且这种映射是可测的，得到的随机变量也有共同的概率空间。
- 直接定义随机变量，而不必依赖于一个底层的抽象概率空间。随机变量 X 的概率空间为 $\{\mathcal{R}, \mathcal{B}(\mathcal{R}), P_X\}$，其中$\mathcal{B}(\mathcal{R})$是通过$\mathcal{R}$上所有开区间来定义的一个 Borel σ 域。通过这个概率架构，我们可以将随机变量归纳到不同的家族，例如高斯、二项分布等，然后以此来对各种物理现象进行建模。

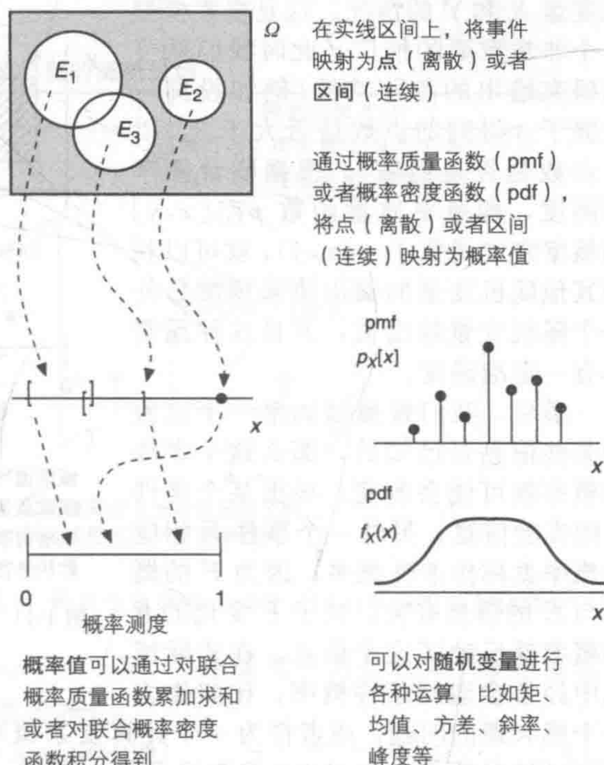


图 1-10　样本空间 Ω 中的事件映射为随机变量 X 的图形化描述。一个随机变量的概率测度可以通过概率质量函数(pmf)(对于离散输出)或者概率密度函数(pdf)(对于连续输出)来定义

从事件到随机变量 X 区间的映射关系如图 1-10 所示。这样做的目的是使用实线区间$\mathcal{R}$来形象描述感兴趣问题中的事件。对于离散随机变量，概率质量函数(pmf)描述了每个可能输出值的概率，我们经常用整数$\mathcal{Z}$的子集来表示。通过对一段区间上相应事件整数值的累加求和，可以得到这个事件的概率值：

$$P(a \leqslant X \leqslant b)=\sum_{x=a}^{b} p_X[x] \tag{1-3}$$

其中 $p_X[x]=P(X=x)$表示概率质量函数，如图 1-10 中的实线圆圈所示。

对于连续随机变量，概率密度函数(pdf)描述了区间$\mathcal{R}$上的概率密度。将概率密度函数在特定区间上进行积分可以求出该事件的概率值：

$$P(a \leqslant X \leqslant b)=\int_a^b f_X(x)\,\mathrm{d}x \tag{1-4}$$

其中 $f_X(x)$表示概率密度函数，该函数是连续的，如图 1-10 所示。此外，离散随机变量和连续随机变量还可以组成混合随机变量。在前面提到过，由于我们对随机变量这个概念比较熟悉，因此在许多情况下就没有必要去考虑这个底层的抽象样本空间，而可以直接对概率质量函数(离散)和概率密度函数(连续)进行操作。

接下来介绍一个重要的拓展，即样本空间 Ω 映射为随机向量的情况：

● 将定义一个随机变量概率空间的方法拓展到定义随机向量，且该向量由两个或多个相关的随机变量构成。在许多情况下，拓展到随机向量是非常有用的，据此我们就可以探讨随机变量的互相关特性了。

图 1-11 形象地描述了两个联合随机变量 X 和 Y 的情况。这是概率模型一个非常重要的推广，此时我们就可以研究输出的各种特性(例如投掷一次骰子，得到的点数是否大于 3，以及点数是否是偶数)。根据联合概率的测度，即概率质量函数 $p_{X,Y}[x,y]$ 和概率密度函数 $f_{X,Y}(x,y)$，就可以利用其他随机变量的输出值来预测另外一个随机变量输出值，并且这种预测具有一定准确度。

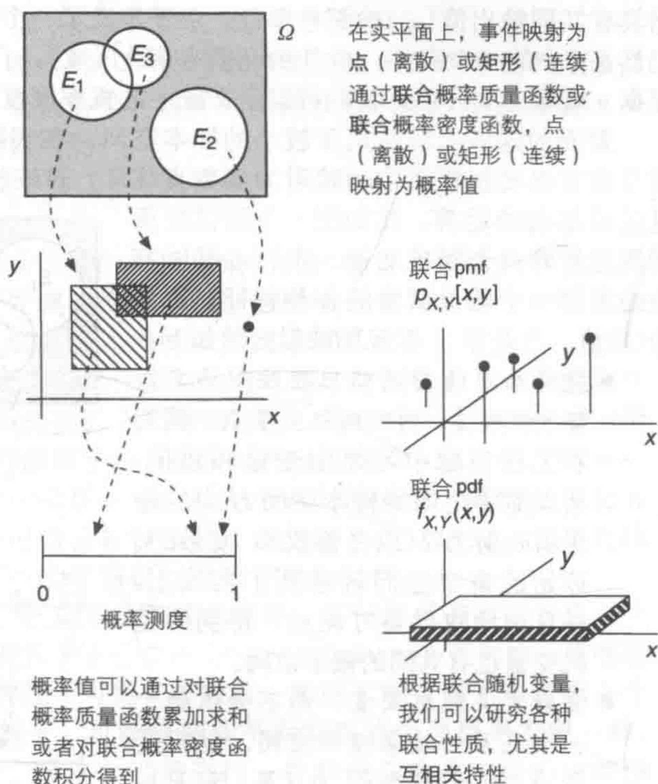


图 1-11 样本空间 Ω 中的事件映射为随机变量 X 和 Y 的图形化描述。可以得到联合矩，我们通常对 X 和 Y 的互相关感兴趣

最后，我们曾提到如果一个试验的某些信息是已知的，那么这个事件的概率就可能会改变。根据某个事件 E 的先验信息，另外一个事件 F 相应的概率就称作条件概率，因为 F 的概率与 E 的信息有关：关于 E 变化的 F 的概率就反映了这个信息。在实际情况中经常会遇到条件概率，比如作为一个感兴趣的问题，或者作为一个计算复杂概率值的方法。在极端情况下，一个试验的"所有"情况都是已知的(如前面提及的，对于一个硬币投掷试验进行无限多次的物理测量)，此时我们就能够依据概率来预测哪个情况会发生。因此，先验概率在概率论中是非常重要的。考虑到这些信息，我们就可以用概率论三个基本公理的拓展来定量研究先验信息对事件概率值的影响。

1.1.3 随机序列和随机过程

随机序列的一个例子就是连续多次的硬币投掷试验。输出结果的每种可能的序列就称作一个现实。例如，如果将一个硬币投掷 10 次，$\{H,H,H,H,H,H,H,H,H,H\}$ 就是一种可能的现实。实际上，这个试验存在 2^{10} 种可能的现实。在此，我们总结一下可以用来描述随机序列和随机过程的方法：

● 定义随机变量的集，并用离散时间 k 来进行索引。$X[k]$代表随机序列，整数 $k\in Z$ 代表离散时间。在某个时刻存在两种可能的输出：如果是离散的，那么就存在有限多个或者可数个输出；或者是连续的，那么输出值是连续的(不可数的)。
● 将随机变量集的概率测度拓展到随机序列。因为已经对联合随机变量定义了 σ 域，那么对于该序列中的任意 N 元随机变量，即使$N\to\infty$也可以定义其相应的概率测度。
● 定义随机变量的集，并用连续时间 t 来进行索引。在实线集合$\mathcal{R}$上定义一个随机过程 $X(t)$，t 表示连续时间。随机序列的情况也是类似的，任意时刻的输出值可能是离散的，也可能是连续的。
● 将随机变量集的概率测度拓展到随机过程。与随机序列不同，随机过程由这些随机变量集组成，但这些随机变量是不可数的。将随机过程的概率测度进行拓展，在拓展时需要满足一致性条件，这将在第 6 章定义。

一共有四种类型的随机序列和随机过程，归纳见表 1-1。与混合型随机变量类似，还可能存在“混合型”的随机序列（或过程），此时一部分输出是连续的，另外一部分是离散的。在后续的讨论中，无论是连续时间过程还是离散时间序列，为简单起见我们都只用“随机过程”来表示。对于只讨论离散时间的情况，我们就会专门用随机序列来表示。对于随机序列，用 $k\in\mathcal{Z}$ 来进行时间上的索引，对于随机过程就用 $t\in\mathcal{R}$ 来进行索引。图 1-12 非常形象地描述了连续时间随机变量映射为离散时间随机变量集（随机序列）。重复类似的试验，可观察其他类型的随机过程。

表 1-1　不同类型的随机过程

随机序列 $X[k]$	随机过程 $X(t)$
离散时间，离散输出（贝努里序列）	连续时间，离散输出（泊松过程）
离散时间，连续输出（高斯序列）	连续时间，连续输出（维纳过程）

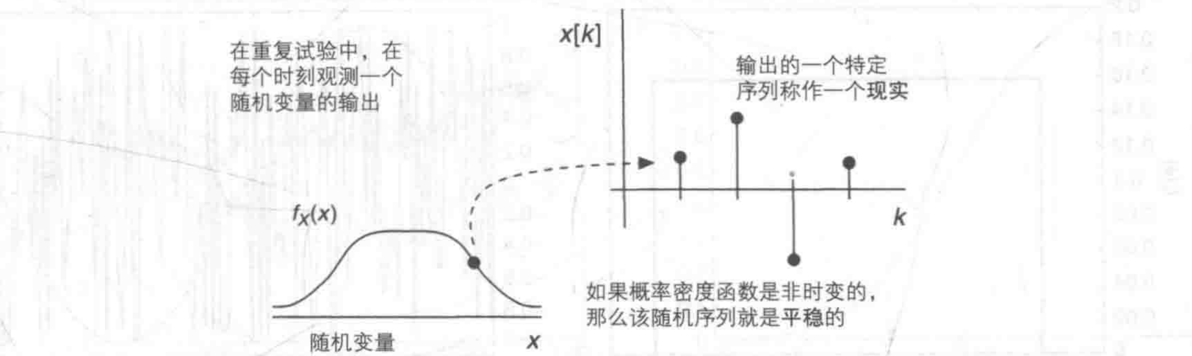


图 1-12　在重复试验中（离散时间，连续输出），随机变量和随机过程之间的联系的图形化表示

例 1-1　考虑一个正弦信号 $x(t)=\sin(\omega t)$，$t\in\mathcal{R}$。一个最简单的随机例子就是随机相位 $\Phi\in[0,2\pi]$，那么就可得到下面这个随机过程：

$$X(t)=\sin(\omega t+\Phi) \tag{1-5}$$

改变 Φ 的取值并不改变信号的波形，只会影响过零点的位置。因为 Φ 是一个连续的数值，存在着无限多可能的输出结果（不可数的），每个结果可以画出一个现实 $x(t)$，得到一簇曲线，如图 1-13a 所示，这簇曲线称作总体。类似地，这个正弦信号的幅度也可以是随机的，$A\in[0,1]$，此时可得，

$$X(t)=A\sin(\omega t) \tag{1-6}$$

或者幅度和相位都是随机的，这种情况的例子如图 1-13b 所示。

$$X(t)=A\sin(\omega t+\Phi) \tag{1-7}$$

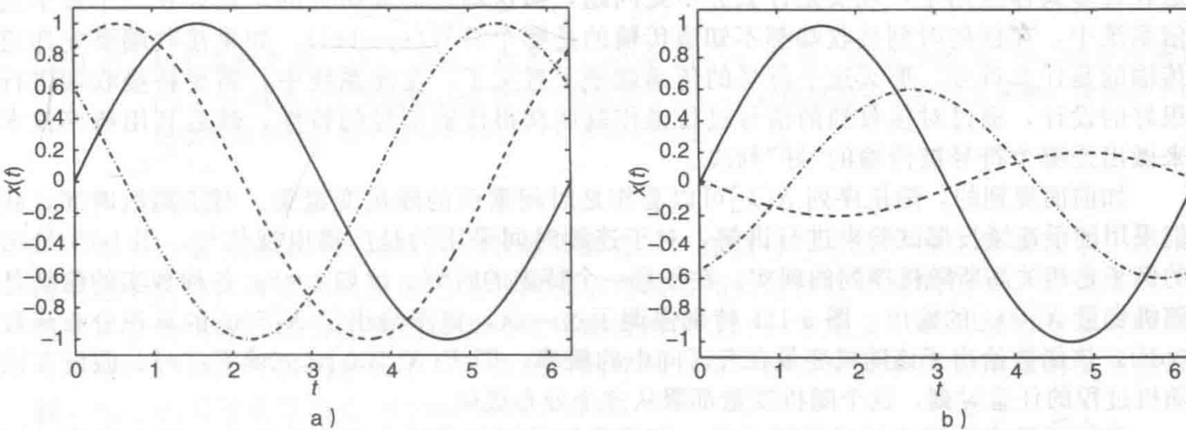


图 1-13　正弦信号 $X(t)=A\sin(\omega t+\Phi)$ 的现实，$\omega=1$。a）相位 Φ 是随机的（$A=1$）；b）相位 Φ 和幅度 A 都是随机的

A和Φ的取值就是随机变量，它们在某个区间上具体取值的概率取决于它们的概率密度函数。相位$\Phi\in[0,2\pi]$的概率密度函数$f_{\Phi}(\phi)$的一个例子如图1-14所示，该分布是均匀的(平坦的)，这表明当$\varepsilon>0$时，对于a的任意取值，$\Phi\in[a,a+\varepsilon]$的概率P都是相同的(确保$[a,a+\varepsilon]$是$[0,2\pi]$的子区间)。随机变量Φ的每个取值都会导致波形的平移，如图1-13a所示，输出值ϕ并不会改变正弦波形的周期。如果相位是时变的，此时随机信号的情况就会更加复杂，因为$\Phi[kT]$本身就是一个随机序列。$\Phi[kT]$幅度的变化意味着相位是随时间t变化的，这等同于正整数与周期T相乘，此时$X(t)$的过零点不再是有规则分布的：

$$X(t)=\sin(\omega t+\Phi[kT]) \tag{1-8}$$

图1-14中的例子表明，即使正弦信号的幅度是区间$[0,1]$中的某个固定值(非随机的)，该信号的现实也是非常易变的，原因就在于相位是时变的。这个时候的信号看上去有点像“噪声”，即使我们非常清楚这只是相位时变的原因。

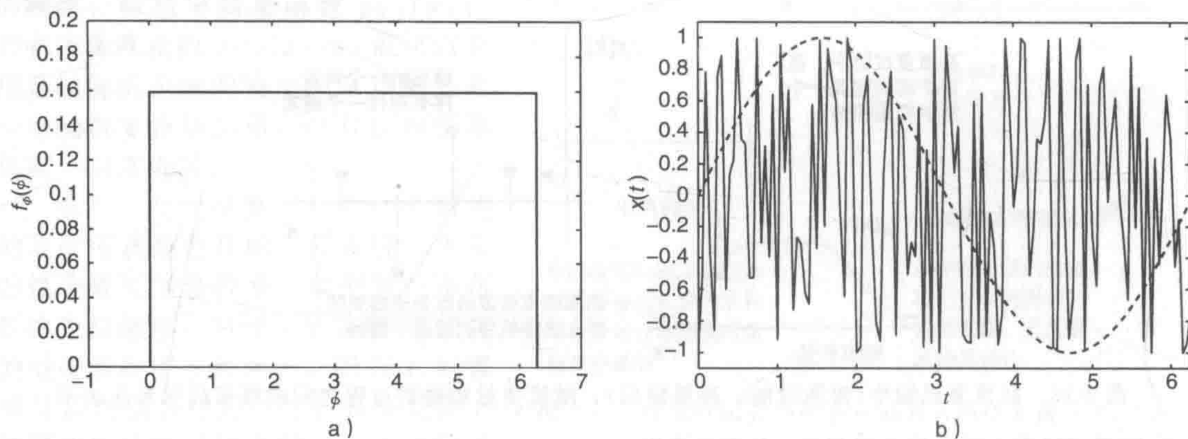


图1-14 正弦信号$X(t)=A\sin(\omega t+\Phi[kT])$，相位$\Phi[kT]$是一个随机序列($A=1,\omega=1,T=0.05$)。a)服从$[0,2\pi]$上均匀分布的随机相位的概率密度函数；b)当相位在不同时刻k独立变化时，正弦信号的现实，$k\in Z^{+}=\{0,1,\cdots\}$

从上面最后一个正弦波形的例子可知，即使是一个简单的随机化，也会导致$x(t)$的现实显得非常不可预测。本书的一个目的就是使我们掌握能够刻画随机信号行为的能力，并且使我们认识随机信号是如何被一个系统处理的。最开始，我们可能会逐个考察样本集中每个现实的行为，但很快我们就会发现这种方法是不现实的(或者是根本没用的)。当然在实际应用中，系统处理的总是样本集里面一个特定的现实，此时我们就可以使用线性系统技术来判断该现实的输出是什么。但由于信号是随机的，我们事先并不知道现实是什么。这在许多具体应用中，现实是什么并不是问题，但这却是需要研究的。比如在一个数字通信系统中，在任何时刻接收端都不知道传输的是哪个符号(symbol)，如果接收端事先知道传输的是什么符号，那么这个符号的传递就毫无意义了。在此系统中，需要将接收端进行很好的设计，通过对接收到的信号进行操作就能获得传输信号的特性，然后利用概率技术来做出是哪个符号被传输的“好”判决。

如前面提到的，随机序列$X[k]$可以看作是时间索引的随机变量集。对于离散时间，我们采用硬币连续投掷试验来进行讲解，对于连续时间采用的是广播电视信号。图1-15a给出的例子是相关高斯随机序列的现实。在任意一个特定的时刻，比如$k=5$，各种现实的值就是随机变量$X[500]$的输出。图5-15b特别强调了这一点，该图给出了$X[500]$的累积分布函数(cdf)，该函数给出了该随机变量在开区间中的概率，即$P(X[500]\leqslant x)\triangleq F_X(x)$。假设在该随机过程的任意时刻，这个随机变量都服从这个分布规律。

我们用概率模型来描述随机过程，即使我们并不知道哪个现实会发生，因此需要理解集(ensemble)的概念。这在各种自然随机过程中是非常有用的，比如每天的气温，它同样

也适用于各种人工随机过程，比如广播信号。在很多实际场合，每个现实的平均值经常是相同的(经常为0)。或者在每个时刻，现实的取值是从高斯分布中“提取”出来的。更重要的是，我们需要通过相关性来定量的描述随机过程在 t_1 时刻，以及后面 t_2 时刻相关联的程度。在许多问题中，比如连续的硬币投掷试验，每次投掷之间并没有任何联系。每次投掷都是独立的，不可能通过已有的结果来预测将来的输出。放大器中的噪声同样也是不可预测的，即使我们可能知道噪声值的大概范围。另一方面，还存在着许多可以被非常准确预测的随机过程。比如在非常短的时间内，一个屋子里的温度是很容易预测的，前后两分钟之间的温度值是高度相关的。图1-15c给出了高斯随机过程自相关函数 $R_{XX}[m]$，该函数是一个双边指数函数的形状。比如在500和1000这两个时刻，该过程的相关程度就是在时延 $m=|k_2-k_1|=500$ 处的取值。随着 m 的增大，相关特性渐渐减少，这是许多随机信号非常典型的特性(不一定是指数衰减的)。

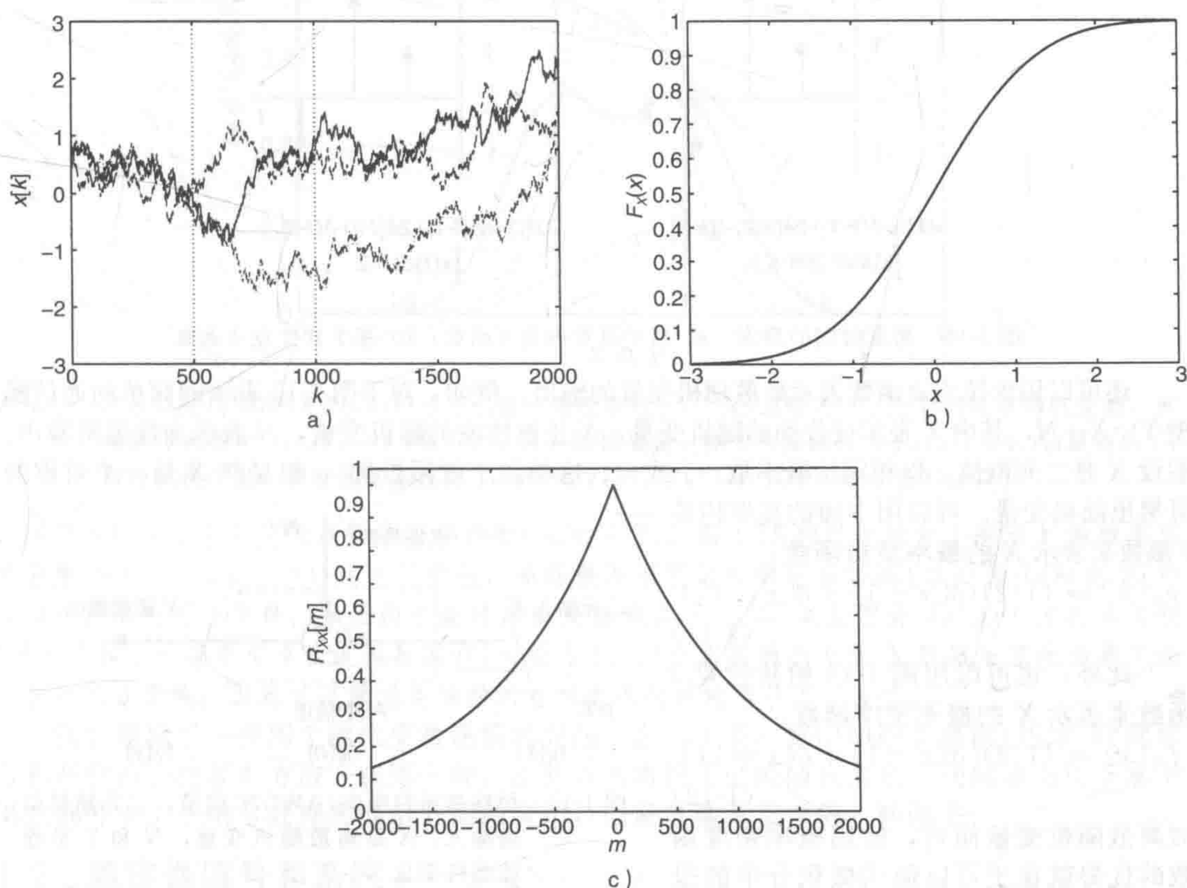


图1-15 相关高斯序列。a) 一个现实的例子；b) 在 $k=500$ 时刻，随机变量 $X[500]$ 的分布函数 $F_X(x)$，这是一个零均值、方差为1的高斯分布；c) 该随机过程的自相关函数，衡量两个时刻输出的差异，横坐标 $m=|k_2-k_1|$ 表示时间差

1.1.4 δ函数

本节讨论 δ 函数，该函数在全书都非常有用。在信号与系统类的工程课程中，经常会遇到这个信号，许多学生很困惑该如何表示离散时间信号和连续时间信号。在这里这个问题同样也会出现，因为随机过程既可以用离散时间(序列)来定义，也可以用连续时间来定义。有两种方法可以用来表示一个离散时间序列。第一种方法经常用于基础的信号处理课程，该方法基于克罗内克(Kronecker)δ函数，

$$\delta[k] \triangleq \begin{cases} 1, & k=0 \\ 0, & \text{其他} \end{cases} \tag{1-9}$$

图 1-16a 给出了一个简单的离散时间现实，用带实心圆的垂直线段表示。只有当$k \in \mathcal{Z}$为整数的时候该过程才存在，在整数之间时间是没有定义的，并且这些现实的积分是无法用黎曼积分来计算的。如果不需要将这个过程拓展到连续时间上，那么这个符号是够用的。但是，有时为了方便，我们想用一个符号既可以描述离散时间序列，又可以描述连续时间过程。此时可以采用狄拉克δ函数，我们在图 1-16b 中用一个现实的示例来进行解释。狄拉克δ函数用一个箭头表示，而且在这些箭头之间时间都是连续、有定义的(严格来讲，狄拉克δ函数不是一个函数，在附录 B 中会讲到它其实是一个广义函数)。比如，我们可以使用卷积和积分运算对其进行操作，因为有下面的等式成立：

$$\int_{-\infty}^{\infty} \delta(t)\mathrm{d}t = 1 \tag{1-10}$$

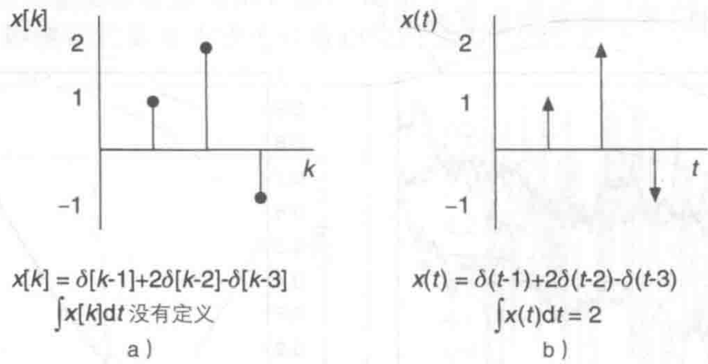


图 1-16　离散时间的现实。a) 基于克罗内克δ函数；b) 基于狄拉克δ函数

还可以用狄拉克δ函数表示离散随机变量的输出。例如，对于图 1-17 表示的简单的通信模型$Y=X+N$，其中Y表示接收到的随机变量，X是被传输的随机变量，N表示加性随机噪声。假设X是二元取值，以相同的概率取-1或$+1$(这类似于硬币投掷)，很显然X是一个对称的贝努里随机变量。可以用下面的克罗内克δ函数来表示X的概率质量函数：

$$p_X[x] = (1/2)(\delta[x-1]+\delta[x+1]) \tag{1-11}$$

此外，也可以用图 1-18 的狄拉克δ函数来表示X的概率密度函数：

$$f_X(x) = (1/2)[\delta(x-1)+\delta(x+1)] \tag{1-12}$$

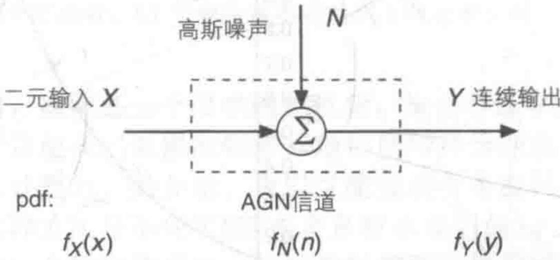


图 1-17　加性高斯白噪声(AWGN)信道，二元随机信道输入。X是离散随机变量，N和Y是连续随机变量

对离散随机变量而言，使用概率密度函数的优势就在于可以使用微积分中的很多方法，因为概率密度函数的定义域是整个实线，而不仅仅是整数。因此，在后面的章节我们都采用概率密度函数来描述离散随机变量的分布情况。同样地，该方法对图 1-17 中的混合随机变量和混合随机过程也是非常方便的。

例 1-2　假设图 1-17 中的噪声服从零均值的高斯分布(也称作正态分布)，其概率密度函数表达式如下(这是一个非常“典型”的曲线)：

$$f_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-n^2/2\sigma^2) \tag{1-13}$$

其中σ^2表示方差。这就是加性高斯噪声(AGN)信道，在通信中这也是最常研究的信道模型，其概率密度函数如图 1-18 所示。根据狄拉克δ函数可知，接收到的随机变量Y的概率密度函数可以由$f_X(x)$(式(1-12))和$f_Y(y)$((式 1-13))相卷积得到(见第 4 章)：

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_X(y-v) f_N(v)\mathrm{d}v \\
&= \frac{1}{2\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} [\delta(y-v-1)+\delta(y-v+1)]\exp(-v^2/2\sigma^2)\mathrm{d}v \\
&= \frac{1}{2\sqrt{2\pi\sigma^2}} [\exp(-(y+1)^2/2\sigma^2)+\exp(-(y-1)^2/2\sigma^2)]
\end{aligned}
\tag{1-14}
$$

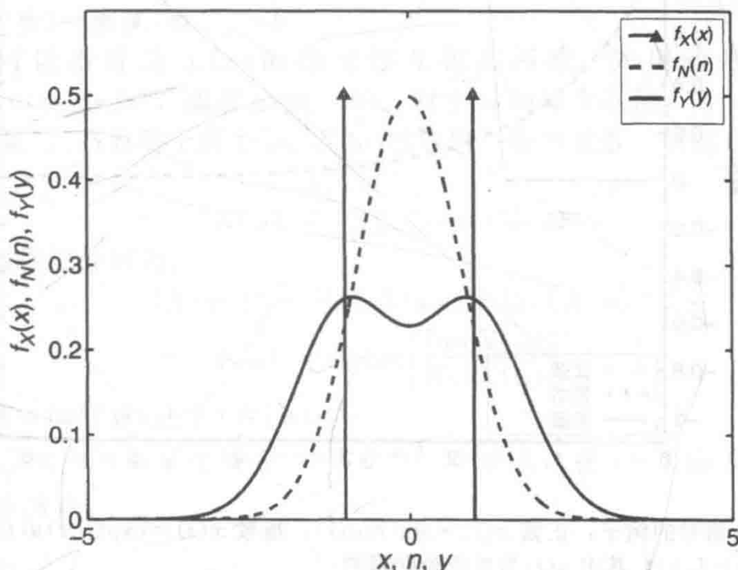


图 1-18　加性高斯白噪声(AWGN)信道上的概率密度函数。(ⅰ)信道输入 X：对称贝努里随机变量，概率密度函数为 $f_X(x)$；(ⅱ)噪声 N：高斯随机变量，概率密度函数为 $f_N(n)$；(ⅲ)信道输出 Y：双峰随机变量，概率密度函数为 $f_Y(y)=f_X(x)*f_N(n)$，其中 $*$ 表示卷积运算

上述推导用到了狄拉克 δ 函数的平移特性(见附录B)。图 1-18 同时也给出了输出 Y 的概率密度函数 $f_Y(y)$。从式(1-14)还可以看出，高斯概率密度函数同时往右边(中心为1)和左边(中心为−1)都进行了平移，并且两个波峰的高度都乘以了1/2，这是因为 $f_Y(y)$ 下方的总面积(两个平移后的高斯概率密度函数的和)必须为1。由于离散随机变量 X 的概率密度函数采用了狄拉克 δ 函数，因此可以通过连续时间卷积来很简单地推导出 $f_Y(y)$。◀

我们预览了一些用于随机变量建模的方法。在1.2节，我们回顾了离散(连续)时间信号和系统的一些基本方法。在第8章，这些方法将用于处理随机过程，同时也会用于本书第三部分中关于信号处理和通信方面的应用，以及其他章节中的一些例子。

1.2　确定性信号和系统

在工程类课程中，学生往往先学习信号与系统在时域中的知识，然后再转到频域学习。图 1-19 就给出了一些确定性信号的例子。

1.2.1　连续时间

非随机信号 $x(t)$ 是连续时间 t 的函数，$t\in\mathcal{R}$。考虑下面的正弦信号

$$x(t)=\sin(\omega t)u(t) \tag{1-15}$$

该信号如图 1-19 所示，其中 $t\in[0,\ 2\pi]$，$u(t)$ 是单位阶跃函数(也称 Heaviside 函数，见附录B)。弧度频率 ω(rad/s)与普通频率 f(Hz)的关系为 $\omega=2\pi f$。对于一个特定的频率 ω，我们可以计算出确定性函数 $x(t)$ 在任意时间 t 的准确数值。图 1-19 同时也给出了其他一些确定性函数的例子。

信号波形还可以在频域进行研究，采用傅里叶变换方法

$$X(\omega)=\int_{-\infty}^{\infty} x(t)\exp(-\mathrm{j}\omega t)\mathrm{d}t \tag{1-16}$$

$$X(f)=\int_{-\infty}^{\infty} x(t)\exp(-\mathrm{j}2\pi ft)\mathrm{d}t \tag{1-17}$$

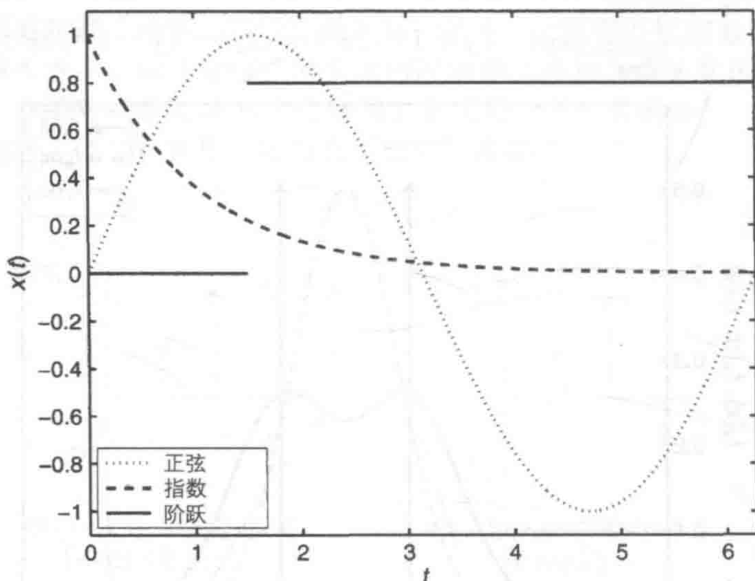


图 1-19　确定性信号的例子：正弦 $x(t)=\sin(t)u(t)$，指数 $x(t)=\exp(-t)u(t)$，阶跃 $x(t)=0.8u(t-1.5)$，其中 $u(t)$为单位阶跃函数

或者也可以采用拉普拉斯变换

$$X(s)=\int_{-\infty}^{\infty} x(t)\exp(-st)\mathrm{d}t \tag{1-18}$$

其中 $s\triangleq\sigma+\mathrm{j}\omega$ 是一个复变量，$\mathrm{j}=\sqrt{-1}$，σ 表示奈培频率(neper frequency)。在附录 C 中对这些变换进行了总结，并归纳了变换的性质和一些有用的变换关系对。(请注意：在这些变换中也采用了大写字母，不要和随机变量混淆在一起了。读者可以根据上下文的描述，以及 $X(\cdot)$的自变量来对这两者进行区分。)这里给出欧拉公式

$$\exp(\mathrm{j}\omega t)=\cos(\omega t)+\mathrm{j}\sin(\omega t) \tag{1-19}$$

$$\exp(-\mathrm{j}\omega t)=\cos(\omega t)-\mathrm{j}\sin(\omega t) \tag{1-20}$$

其中 $\cos(\omega t)$和 $\sin(\omega t)$分别表示 $\exp(\mathrm{j}\omega t)$的实部和虚部。拉普拉斯变换 $X(s)$是 σ 和 ω 的函数，一般都在 s 平面上通过零极点来表示，零点表示 $X(s)$为 0 的位置(用○表示)，极点表示 $X(s)$取无穷大的位置(用×表示)。

对应的逆变换公式为：

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\exp(\mathrm{j}\omega t)\mathrm{d}\omega \tag{1-21}$$

$$=\int_{-\infty}^{\infty} X(f)\exp(2\pi ft)\mathrm{d}f \tag{1-22}$$

$$=\oint_C X(s)\exp(st)\mathrm{d}s \tag{1-23}$$

其中上面最后一个等式表示在复 s 平面上收敛域(ROC)范围内的一个围线积分。此外，通过变换 $\omega=2\pi f$，将 $\mathrm{d}\omega=2\pi\mathrm{d}f$ 带入第一个积分表达式，可以推导出第二个积分表达式。

例 1-3 考虑如下的信号：

$$x(t)=\cos(2\pi f_0 t) \tag{1-24}$$

其中 $t\in\mathcal{R}$，频率 f_0 的单位是 Hz。利用狄拉克 δ 函数，这个信号频谱的数学表达式为：

$$X(f)=(1/2)[\delta(f-f_0)+\delta(f+f_0)] \tag{1-25}$$

将 $x(t)$ 和 $X(f)$ 写成弧度频率 ω 的函数：

$$x(t)=\cos(\omega_0 t) \tag{1-26}$$

$$X(\omega)=\pi[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)] \tag{1-27}$$

正弦信号的频率是Hz，表示在一秒钟内该信号一个周期（一个循环）发生的次数。弧度频率是一种角度单位，正弦信号一个周期的变化就是 2π 弧度。弧度 ω 的转换关系为：（弧度/循环）（循环/秒）＝弧度/秒。◀

这些变换可以看作是 $x(t)$ 的确定性互相关函数，采用的是复指数表达方式 $\exp(-j\omega t)$，$\exp(-j2\pi ft)$，或者 $\exp(-st)$。对于 ω 的每个取值，$X(\omega)$ 表示 $x(t)$ 中该频率的强度或“数量”。因为对于每个 ω，$X(\omega)$ 的取值一般为复数，因此它可以表示成如下的极坐标形式：

$$X(\omega)=|X(\omega)|\exp(j\phi(\omega)) \tag{1-28}$$

其中平方幅度和相位分别为：

$$|X(\omega)|^2=\mathrm{Re}^2[X(\omega)]+\mathrm{Im}^2[X(\omega)] \tag{1-29}$$

$$\phi(\omega)=\arctan\left[\frac{\mathrm{Im}(X(\omega))}{\mathrm{Re}(X(\omega))}\right] \tag{1-30}$$

$x(t)$ 中频率分量 ω 的数量（强度）为 $|X(\omega)|$。

例 1-4 下面的拉普拉斯变换在 $s=1$ 处有一个极点，在 $s=\pm j\omega$ 处有一对复零点，如图1-20的 s 平面所示：

$$X(s)=\frac{s^2+1}{s-1} \tag{1-31}$$

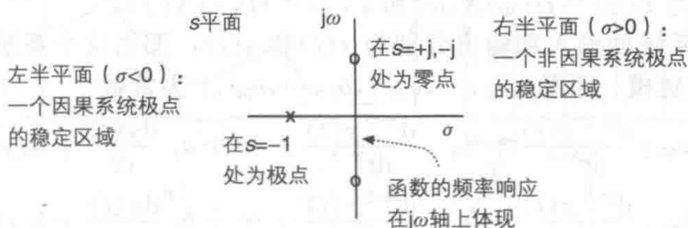


图1-20　例1-4中拉普拉斯变换 $X(s)=(s^2+1)/(s-1)$ 在 s 平面上的零极点

极点体现了 $x(t)$ 的很多特性，比如有界性（$\sigma<0$）、衰减速率（看 σ 与0的接近程度），以及判断有无正弦分量（当极点不在实轴上时存在正弦分量）。将 $X(s)$ 在 $j\omega$ 上投影就可以得到 $x(t)$ 的傅里叶变换，式(1-31)的傅里叶变换为：

$$X(\omega)=\frac{1-\omega^2}{j\omega-1}=\frac{\omega^2-1+j\omega(\omega^2-1)}{\omega^2+1}=a+jb \tag{1-32}$$

其中 $a=(\omega^2-1)/(\omega^2+1)$，$b=\omega(\omega^2-1)/(\omega^2+1)$。$X(\omega)$ 的平方幅度和相位分别为：

$$|X(\omega)|^2=a^2+b^2=\frac{(\omega^2-1)^2}{\omega^2+1} \tag{1-33}$$

$$\arg[X(\omega)]=\arctan(b/a)=\arctan(\omega) \tag{1-34}$$

以 ω 为横坐标，图1-21分别绘出了平方幅度和相位。$X(\omega)$ 给出了信号频率分量的许多信息，因此也被称作频谱。一般将平方幅度以分贝(dB)为单位给出：

$$10\lg(|X(\omega)|^2) \tag{1-35}$$

可以看出，图形中“开口”由零点 $z=\pm j$ 的位置决定，相位在 $\omega=0$ 附近几乎为线性的。◀

工程课程中讲到的系统往往都是线性时不变(LTI)滤波器，冲激响应函数为 $h(t)$。当输入为 $x(t)$ 时，通过下面的卷积积分可求出该系统的输出为：

$$y(t)=\int_0^{\infty}h(t-\tau)x(\tau)d\tau=\int_0^{\infty}x(t-\tau)h(\tau)d\tau \tag{1-36}$$

可以看出，$x(t)$和$h(t)$是对称的：$y(t)=h(t)*x(t)=x(t)*h(t)$。系统的线性性质表明结果是可叠加的，系统的时不变特性表明冲激响应不随时间的平移而发生变化，因此上面的$h(t-\tau)$可以写成一个更一般的表达式$h(t,\tau)$。$h(t)$的拉普拉斯变换就是系统的传递函数$H(s)$，$h(t)$的傅里叶变换$H(\omega)$就是对应的频率相应。在s域中线性时不变系统的输出为：

$$Y(s)=H(s)X(s) \tag{1-37}$$

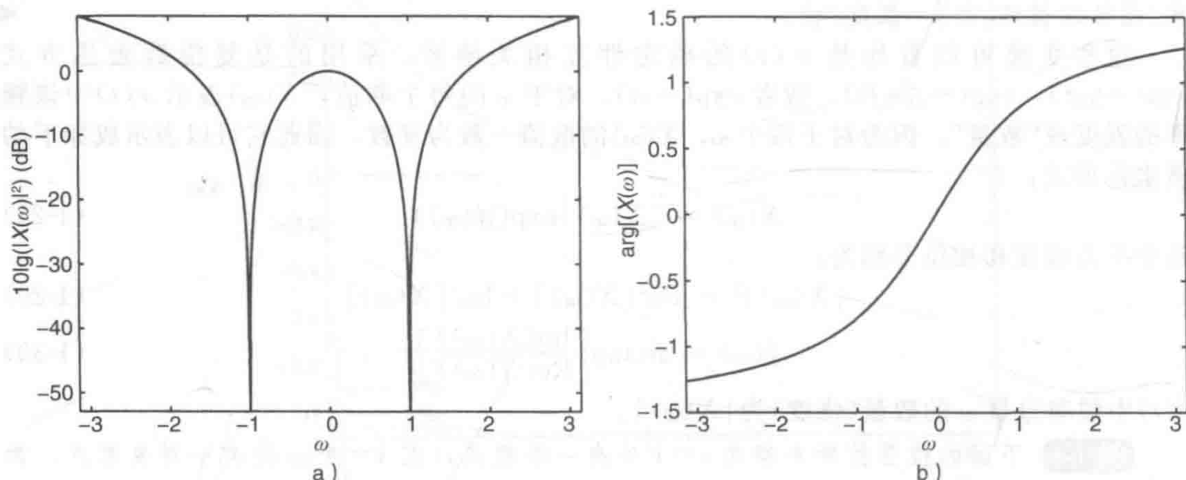


图 1-21 例 1-4 中的傅里叶变换 $X(\omega)$由 $X(s)=(s^2+1)/(s-1)$得出。a)傅里叶变换的平方幅度（单位为 dB）；b)傅里叶变换的相位

傅里叶域中的输出为$Y(\omega)=H(\omega)X(\omega)$和$Y(f)=H(f)X(f)$。

如果一个线性系统的输入和输出分别为$x(t)$和$y(t)$，那么这个系统就可以用下面的线性差分方程(DE)来建模，系数$\{a_0,\cdots,a_{N-1},b_0,\cdots,b_{M-1}\}$为常数

$$\begin{aligned}&a_{N-1}\frac{\mathrm{d}^{N-1}y(t)}{\mathrm{d}t^{N-1}}+a_{N-2}\frac{\mathrm{d}^{N-2}y(t)}{\mathrm{d}t^{N-2}}+\cdots+a_1\frac{\mathrm{d}y(t)}{\mathrm{d}t}+a_0y(t)\\&=b_{M-1}\frac{\mathrm{d}^{M-1}x(t)}{\mathrm{d}t^{M-1}}+b_{M-2}\frac{\mathrm{d}^{M-2}x(t)}{\mathrm{d}t^{M-2}}+\cdots+b_1\frac{\mathrm{d}x(t)}{\mathrm{d}t}+b_0x(t)\end{aligned} \tag{1-38}$$

或者也可以写成如下更紧凑的表达式

$$\sum_{n=0}^{N-1}a_ny^{(n)}(t)=\sum_{m=0}^{M-1}b_mx^{(m)}(t) \tag{1-39}$$

上标表示求导的阶数。假设初始状态为0，传递函数就可以写成两个多项式相除的形式

$$H(s)=\frac{Y(s)}{X(s)}=\frac{\sum_{m=0}^{M-1}b_ms^m}{\sum_{n=0}^{N-1}a_ns^n} \tag{1-40}$$

将上式进行拉普拉斯逆变换就可以求出$h(t)$（一般情况下$a_0=1$）。这种有理函数是很重要的，因为通过部分分式展开法(PFE)可以很简单地求出逆变换。

我们在时域和频域对确定性信号进行了描述，也非常有兴趣将这种思路拓展到随机信号，但我们并不打算研究随机系统。但在后面的第11～13章有几个例外，比如卡尔曼滤波器、自适应滤波器和盲均衡器等，为了满足某些优化准则这些滤波器是时变的。一般情况下，我们假设系统是固定的，用一个恒系数的线性差分方程来对该系统进行建模，其冲激响应函数为$h(t)$。当然了，许多实际系统往往是时变的，此时线性差分方程只是系统动力学的一个近似模型。实际上，许多系统是非线性的，或者说这些系统是线性分量和非线性分量的组合。接下来，我们考虑一个可能是最简单的非线性系统，即无记忆系统。无记

忆系统不存储输入或输出的任何过去的信息，它当前时刻的输出只取决于当前时刻的输入。这种系统可以用如下的方程来表示，

$$y(t) = g(x(t)) \tag{1-41}$$

其中 $g(\cdot)$ 是线性时不变的，$y(t)$ 在每个时刻的取值就是这个函数输入 $x(t)$ 的结果。图 1-22 归纳了上面讨论的两种系统。

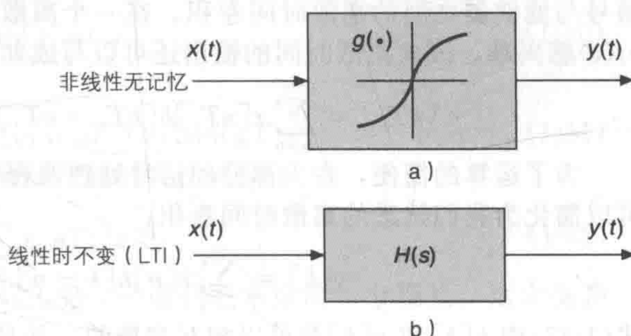


图 1-22 本书中讨论的两种系统。a)非线性无记忆映射 $g(\cdot)$；b) 线性时不变系统，冲激响应函数为 $h(t)$，传递函数为 $H(s)$

1.2.2 离散时间

离散时间信号可以用克罗内克 δ 函数或者狄拉克 δ 函数来表示。例如，如果对一个连续时间信号 $x(t)$ 进行采样得到离散时间信号，很显然这个采样过程可以用如下的乘法表示：

$$x_s(t) = x(t)s(t) \tag{1-42}$$

其中 $x_s(t)$ 表示采样后的信号：

$$s(t) = \sum_{n=0}^{\infty} \delta(t - nT_s) \tag{1-43}$$

式中，$s(t)$ 表示采样信号，T_s 是采样周期，采样频率 $f_s = 1/T_s$。图 1-23 描述了一个均匀采样的过程。通过采用狄拉克 δ 函数，$x_s(t)$ 在所有的 t 上都有定义，即使 $x_s(t)$ 只在一些离散采样时刻取值为非 0。在此假设满足奈奎斯特采样定理：采样率 f_s 大于 $x(t)$ 所有分量中最高频率的两倍，这样就不会发生混叠现象(也就是说频率响应平移后不会发生重叠)。

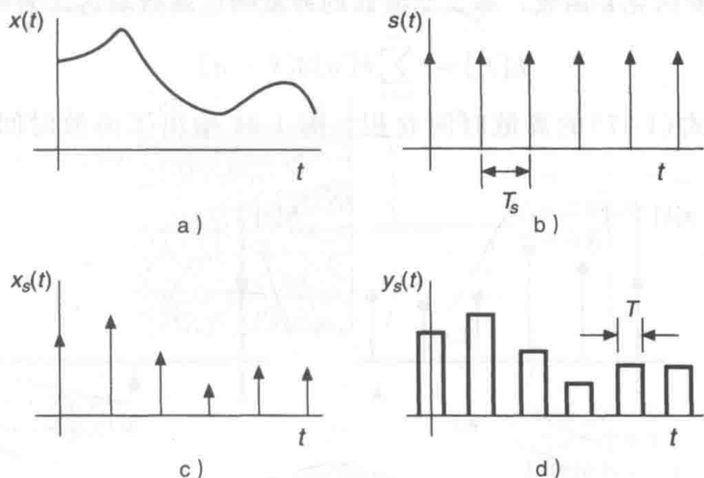


图 1-23 对一个确定性信号进行采样。a) 连续时间信号 $x(t)$；b) 根据狄拉克 δ 函数定义的采样信号 $s(t)$，**采样周期 T_s 是均匀的**；c) 采样后的信号 $x_s(t)$；d) 滤波后的信号 $y_s(t) = x_s(t) * h(t)$，其中 $h(t)$ 是一个矩形冲激响应函数，$T < T_s$

根据狄拉克 δ 函数的采样特性可知，采样后的信号波形可以表示为如下形式：

$$x_s(t) = \sum_{n=0}^{\infty} x(nT_s)\delta(t - nT_s) \tag{1-44}$$

从图 1-23 可以看出，$x_s(t)$ 是“一长串”狄拉克 δ 函数，每个冲激的强度由 $x(t)$ 在该采样时刻的取值决定。前面提到过，因为 $x_s(t)$ 在所有的 t 上都有定义，使用狄拉克 δ 函数的优势在于可以很轻松地进行积分运算。例如，假设用一个连续时间滤波器 $h(t)$ 来对 $x_s(t)$ 进行滤波，通过卷积积分可得：

$$y_s(t) = \int_0^{\infty} x_s(\tau)h(t-\tau)\mathrm{d}\tau = \sum_{n=0}^{\infty} x(nT_s)\int_0^{\infty} \delta(\tau - nT_s)h(t-\tau)\mathrm{d}\tau$$

$$= \sum_{n=0}^{\infty} x(nT_s)h(t-nT_s) \tag{1-45}$$

上式用到了狄拉克 δ 函数的平移特性。从图 1-23 可以看出，$y_s(t)$仍然是一个连续时间函数，假设矩形函数 $h(t)$的宽度 T 不大于采样周期 T_s。式(1-45)中的最终表达式就是输入信号与滤波器之间的离散时间卷积。在一个离散时间系统中，我们只对整数倍 T_s 时刻的 $y_s(t)$感兴趣，因此离散时间的输出还可以写成如下形式：

$$y_s[kT_s] = \sum_{n=0}^{\infty} x[nT_s]h[kT_s - nT_s] = \sum_{n=0}^{\infty} x[nT_s]h[(k-n)T_s] \tag{1-46}$$

为了运算的简便，在大部分的信号处理课程中一般都假设 $T_s=1$，因此上面的表达式可以简化为我们熟悉的离散时间卷积：

$$y[k] = \sum_{n=0}^{\infty} x[n]h[k-n] = \sum_{n=0}^{\infty} h[n]x[k-n] \tag{1-47}$$

式(1-47)中 $h[k]$和 $x[k]$是可以相互交换的，并且 $y_s[k]$的下标也省略了，因为这里全都是离散时间变量(在此不考虑对连续时间信号的采样过程)。如果只考虑离散时间信号和系统，那么就应该使用克罗内克 δ 函数，而不再使用狄拉克 δ 函数，此时式(1-44)写为：

$$x[k] = \sum_{n=0}^{\infty} x[n]\delta[k-n] \tag{1-48}$$

式(1-48)的求和表达式看上去有些复杂，因为等号的两端都出现了 $x[\cdot]$，在等号左边代入一个具体的 k，等号右端就会取到某个取值(比如代入 $k=1$，因为 δ 函数的缘故，那么在求和表达式中就只有 $n=1$ 这一项出现)。图 1-24 给出了 $x_s(t)$的离散时间形式 $x[k]$。如果采用克罗内克 δ 函数，那么滤波器的冲激响应函数表达式为如下形式：

$$h[k] = \sum_{n=0}^{\infty} h[n]\delta[k-n] \tag{1-49}$$

滤波器的输出就是式(1-47)的离散时间卷积。图 1-24 给出了离散时间滤波器以及输出 $y[k]$。

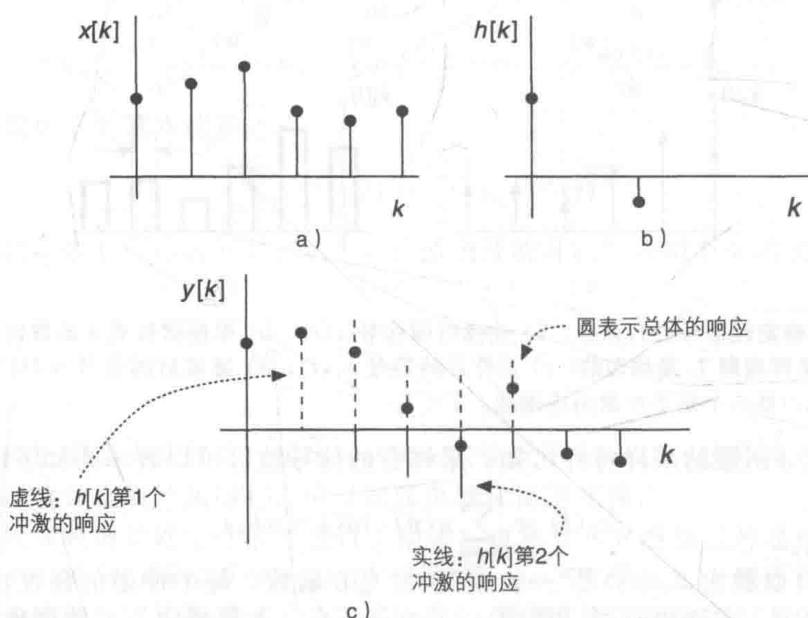


图 1-24　离散时间信号和滤波器。a) 输入信号 $x[k]$；b) 滤波器的冲激响应函数 $h[k]$；c) 滤波器输出 $y[k]=x[k]*h[k]$

采样后信号的拉普拉斯变换为：

$$
\begin{aligned}
X_s(s) &= \int_0^{\infty} x_s(t)\exp(-st)\,\mathrm{d}t \\
&= \sum_{n=0}^{\infty} x[nT_s]\int_0^{\infty} \delta(t-nT_s)\exp(-st)\,\mathrm{d}t \\
&= \sum_{n=0}^{\infty} x[nT_s]\exp[-nT_s s]
\end{aligned}
\tag{1-50}
$$

可以看出式(1-51)的周期为 T_s

$$
\exp(sT_s) = \exp(\sigma T_s)[\cos(\omega T_s) + \mathrm{j}\sin(\omega T_s)] \tag{1-51}
$$

定义 $z \triangleq \exp(s)$，通过拉普拉斯变换直接得到 z 变换的表达式如下：

$$
X(z) = \sum_{n=0}^{\infty} x[nT_s] z^{-nT_s} \tag{1-52}
$$

从图 1-25 可以看出，映射 $z=\exp(s)$实际上是 s 平面到 z 平面的一个变换。这个变换具有以下一些特性：

- z 平面上的实轴为$\exp(\sigma T_s)\cos(\omega T_s)$(式(1-51)的第 1 项)。$z$ 平面的虚轴为$\exp(\sigma T_s)\sin(\omega T_s)$(式(1-51)的第 2 项)。$\sigma=0$ 时表示单位圆，在 z 平面上的表达式为：

$$
\cos(\omega T_s) + \mathrm{j}\sin(\omega T_s) = \exp(\mathrm{j}\omega T_s) \tag{1-53}
$$

- s 平面上的左半部分($\sigma<0$)映射到单位圆内(因为 $\exp(\sigma T_s)<1$)。s 平面上的右半部分($\sigma>0$)映射到单位圆外(因为 $\exp(\sigma T_s)>1$)。s 平面左半平面上的一系列水平“线段”映射为 z 平面上的“扇形区”。当 $\sigma=-\infty$时。所有的 ω 映射为原点 $z=0$。
- $\mathrm{j}\omega$ 轴上的区间 $[-\omega_s/2, \omega_s/2]$ 映射为 z 平面上的单位圆。当 $k\in\mathbb{Z}$时，$\mathrm{j}\omega$ 轴上所有类似于 $[-(k+1)\omega_s/2, -k\omega_s/2]$, $[k\omega_s/2, (k+1)\omega_s/2]$ 的区间都映射为 z 平面上的单位圆。

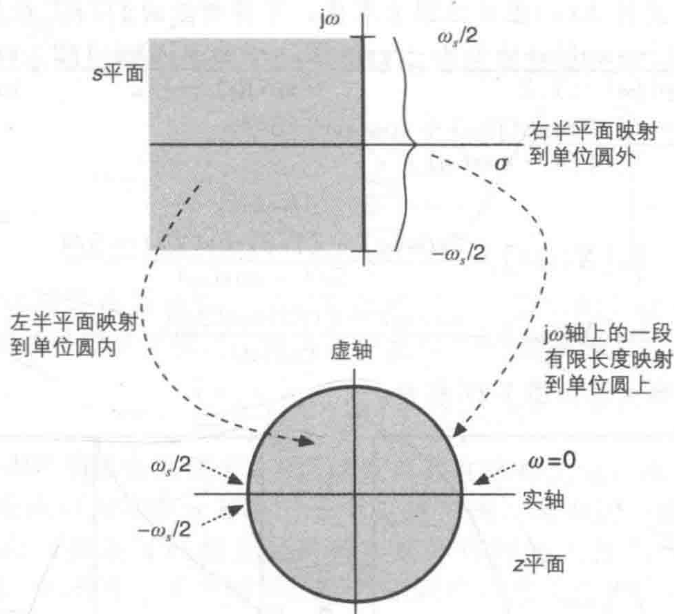


图 1-25 s 平面到 z 平面的映射

在式(1-52)中代入 $\sigma=0$，得到 $r=1$ 和 $z=\exp(\mathrm{j}\omega)$，可得该离散时间信号的频率响应为：

$$
X(\mathrm{j}\omega) = \sum_{k=0}^{\infty} x[k]\exp(-\mathrm{j}\omega k) \tag{1-54}
$$

其中 $T_s=1$。我们在 $X(\mathrm{j}\omega)$的表达式中用了变量 j，这样离散时间傅里叶变换(DTFT)就不会同连续时间傅里叶变换 $X(\omega)$相互混淆。从图 1-25 以及上面的讨论可知，$X(\mathrm{j}\omega)$“环绕”

着单位圆，当 $\omega=\pm\pi$ 时，$X(j\omega)$映射到单位圆最左端的点(设 $T_s=1$)。利用欧拉公式可以很容易验证这个结论：

$$z=r\exp(j\omega)=r\cos(\omega)+jr\sin(\omega) \tag{1-55}$$

当 $z=-1$(单位圆最左边的那个点)和 $r=1$ 时，正弦函数取值为 0，余弦函数取值为 -1。对于所有非 0 整数倍的 $\pm\pi$，这个结论始终成立。当$|\omega|$超过 π 时，频率响应重复出现。这种重复特性是采样的结果，采样过程使得原始的频率响应进行了多次平移，以整数倍采样率($f_s=1/T_s$)为单位进行平移。当 $T_s\neq1$ 时，单位圆上最左边的点为$\pm f_s/2=1/2T_s$(采样频率的一半)。对于弧度频率，单位圆上最左边的点为$\pm\omega_s/2=\pm\pi f_s$，如图 1-25 所示。因为频率响应是重复出现的，因此一般只画出$|\omega|\leqslant\pi$ 的结果($T_s=1$)。

例 1-5 图 1-26 给出的是下面这个 z 变换的零极点位置示意图。

$$X(z)=\frac{z^2-z+1/2}{z-1/2} \tag{1-56}$$

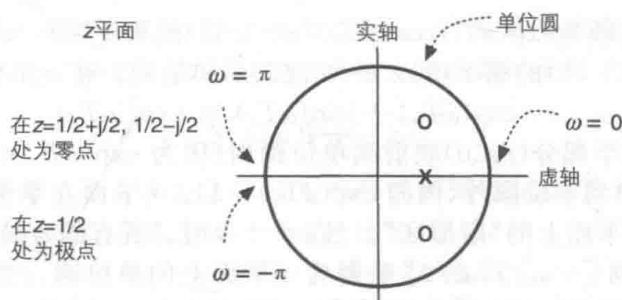


图 1-26　例 1-5 中 $X(z)$的零极点位置示意图

代入 $z=\exp(j\omega)$，并且将 $X(z)$在单位圆上取值，可得相应的 DTFT 结果：

$$\begin{aligned}X(j\omega)&=\frac{\exp(j2\omega)-\exp(j\omega)+1/2}{\exp(j\omega)-1/2}=\frac{\exp(j2\omega)-\exp(j\omega)+1/2}{\exp(j\omega)-1/2}\cdot\frac{\exp(-j\omega)-1/2}{\exp(-j\omega)-1/2}\\&=\frac{\exp(j\omega)-(1/2)\exp(j2\omega)+\cos(\omega)-5/4}{5/4-\cos(\omega)}\end{aligned} \tag{1-57}$$

实部和虚部分别为：

$$\mathrm{Re}[X(j\omega)]=\frac{2\cos(\omega)-(1/2)\cos(2\omega)-5/4}{5/4-\cos(\omega)} \tag{1-58}$$

$$\mathrm{Im}[X(j\omega)]=\frac{\sin(\omega)-(1/2)\sin(2\omega)}{5/4-\cos(\omega)} \tag{1-59}$$

其幅频特性和相频特性曲线如图 1-27 所示。

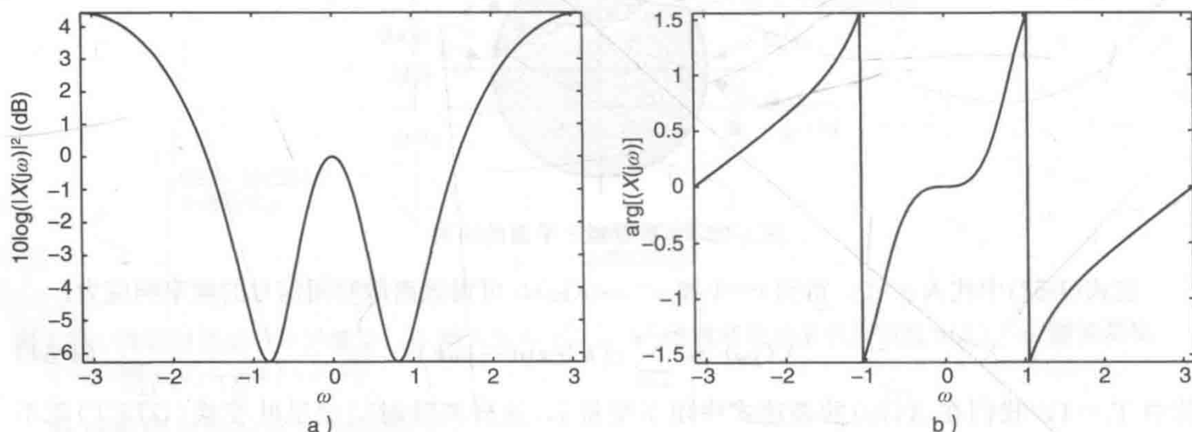


图 1-27　例 1-5 中 $X(z)$的 DTFT 结果。a)幅度平方(dB)；b)相位

1.2.3　离散时间滤波器

将式(1-47)中的卷积结果表示为滤波器的输出，在 z 域上的结果为：

$$Y(z)=H(z)X(z) \tag{1-60}$$

其中，下式就是传递函数：

$$H(z)=\sum_{k=0}^{\infty}h[k]z^{-k} \tag{1-61}$$

同理可得输出的频率响应为：

$$Y(\mathrm{j}\omega)=H(\mathrm{j}\omega)X(\mathrm{j}\omega) \tag{1-62}$$

如果式(1-47)和式(1-61)中的求和项都是有限的，那么这就是一个有限长度冲激响应(FIR)的滤波器。如果求和限为 $\{0,M-1\}$，传递函数如下：

$$H_{\mathrm{FIR}}(z)=\sum_{m=0}^{M-1}b_m z^{-m} \tag{1-63}$$

此时式(1-47)变为：

$$y[k]=\sum_{m=0}^{M-1}b_m x[k-m] \tag{1-64}$$

其中 $\{b_m\}$ 表示 FIR 滤波器的各阶系数。因为求和下限大于等于 0，因此这个滤波器是因果的。图 1-28 给出了一个 FIR 滤波器的结构框图。它包含一个多抽头的延迟线(TDL)结构，可用于存储输入值 $x[k]$ 的时延数据，比如 $\{x[k-1],\cdots,x[k-M+1]\}$。在图中用 z^{-1} 来表示时延运算符，表示输入 $x[k]$，输出 $x[k-1]$。

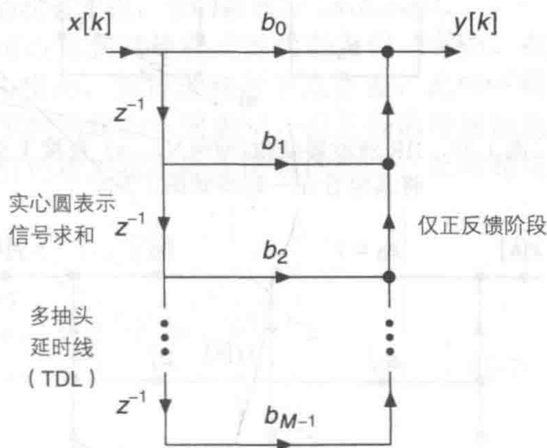


图 1-28　FIR 滤波器的结构

如果式(1-47)的求和上限为无穷大，那么这个滤波器就是无限冲激响应(IIR)的。此时传递函数为：

$$H_{\mathrm{IIR}}(z)=\frac{\sum_{m=0}^{M-1}b_m z^{-m}}{\sum_{n=0}^{N-1}a_n z^{-n}} \tag{1-65}$$

上式给出的是离散时间的传递函数，与之对比式(1-40)给出的是连续时间的传递函数。式(1-65)表示的滤波器可以从下面这个线性差分方程(DE)推导出来，在此假设 $a_0=1$

$$y[k]=\sum_{m=0}^{M-1}b_m x[k-m]+\sum_{n=1}^{N-1}a_n y[k-n] \tag{1-66}$$

图 1-29 给出了一些 IIR 滤波器的例子。IIR 滤波器存在多种类型，图 1-29a 所示的结构称为直接Ⅰ型，它就是式(1-66)差分方程的一个直接实现。因为这个滤波器是线性时不变的，它的前向和后向网格是可以相互交换的，这样可得第二种类型的 IIR 滤波器，如图 1-29b所示。在这个结构中，两个时延信号是相同的，因此这两个时延信号线可以整合为一条，得到直接Ⅱ型，如图 1-30a 所示。这种结构具有最少数目的延迟单元，延迟单元的数目为 $\max(M-1,N-1)$。

1.2.4　状态空间的实现

最后，我们简要介绍一下状态空间的实现。利用状态空间可以很方便地描述一个线性系统，也可以用于第 11 章将要介绍的卡尔曼滤波器。式(1-66)原为一个高阶的 IIR 差分方程，将其改写为一阶向量差分方程就可得到状态空间的实现。通过将滤波器的内部信号定义为状态 $\{x_n[k]\}$，可得如下的状态向量：

$$\boldsymbol{x}[k] \triangleq [x_1[k], \cdots, x_{N-1}[k]]^{\mathrm{T}} \tag{1-67}$$

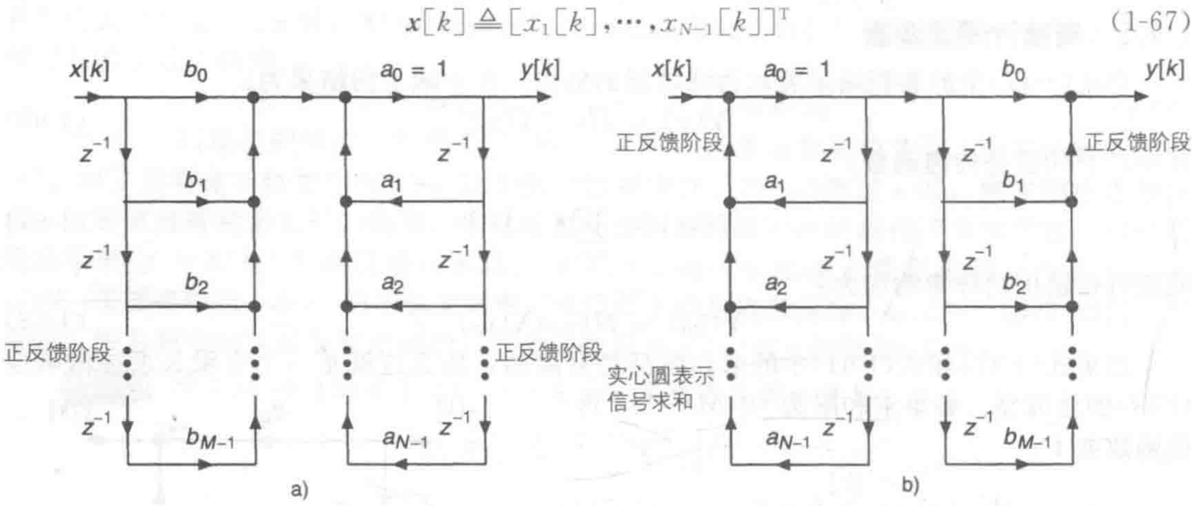


图 1-29　IIR 滤波器结构($M=N$)。a) 直接Ⅰ型；b) 直接Ⅱ型，两个降阶时延线上的信号是相同的，将其组合在一起得到图 1-30a

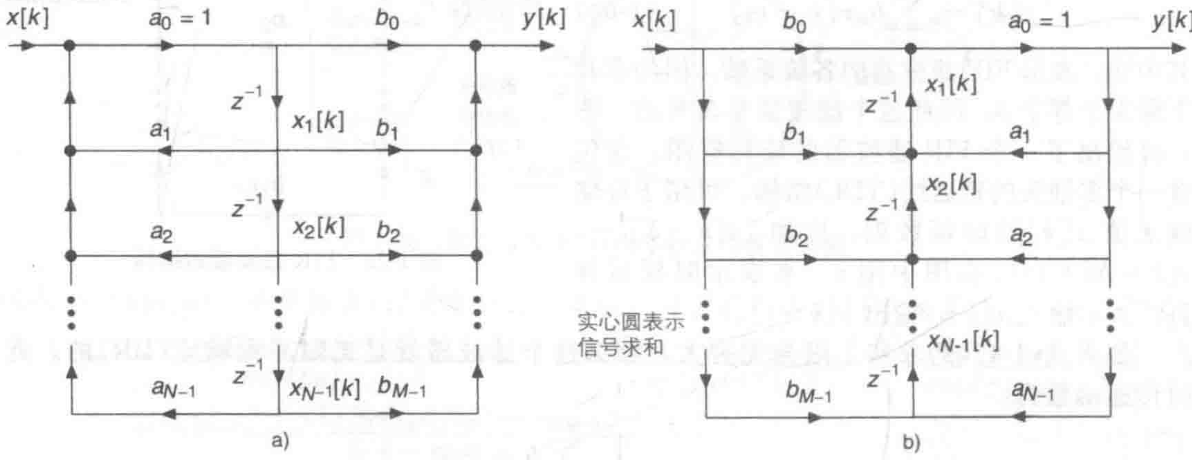


图 1-30　IIR 滤波器结构($M=N$)。a) 定义了状态信号的直接Ⅱ型(将图 1-29b 中的两个降阶时延线组合在一起)；b) 另外一种配置，将升阶时延线组合在一起，此时状态信号的定义不同

图 1-30a 中，为简单起见我们假设 $M=N$。状态数目为 $\max(M-1, N-1)$。在直接Ⅱ型结构中的 $N-1$ 种状态其实就是时延线上的信号，因此我们可以写出下面的系统方程：

$$\begin{aligned} x_1[k+1] &= \sum_{n=1}^{N-1} a_n x_n[k] + x[k] \\ x_2[k+1] &= x_1[k] \\ &\vdots \\ x_{N-1}[k+1] &= x_{N-2}[k] \end{aligned} \tag{1-68}$$

根据状态向量的定义，可得下面的递归向量表达式：

$$\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{b}x[k] \tag{1-69}$$

其中，

$$\boldsymbol{A} \triangleq \begin{bmatrix} a_1 & a_2 & \cdots & & a_{N-1} \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{b} \triangleq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{1-70}$$

差分方程还可以用下面的状态来进行表示(仍然假设 $M=N$)：

$$y[k]=\sum_{m=1}^{M-1}b_m x_m[k]+b_0\sum_{n=1}^{N-1}a_n x_n[k]+b_0 x[k] \tag{1-71}$$

$$=\sum_{n=1}^{N-1}(b_n+b_0 a_n)x_n[k]+b_0 x[k] \tag{1-72}$$

此时的输出结果为：

$$y[k]=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}[k]+\mathrm{d}x[k] \tag{1-73}$$

其中，

$$\boldsymbol{c}\triangleq[b_1+b_0a_1,\cdots,b_{N-1}+b_0a_{N-1}]^{\mathrm{T}},\quad d=b_0 \tag{1-74}$$

式(1-69)和式(1-73)共同组成了直接Ⅱ型滤波器的状态方程，它的系数为 $\{\boldsymbol{A},\boldsymbol{b},\boldsymbol{c},d\}$。

将直接Ⅱ型滤波器的现实进行重新整理，还可以得到其他形式的状态方程。例如，在图 1-30b 中，交换 $x[k]$和 $y[k]$，反转所有的箭头指向，交换求和与节点位置，此时可得水平框图的对偶现实。图中所示的两个现实具有不同的状态空间表示，但它们的传递函数都是相同的(零点极点相同)。在图 1-30b 中，我们仍然把状态作为时延的输出，此时相应的状态方程为($M=N$)：

$$\begin{aligned}x_1[k+1]&=b_1x[k]+a_1y[k]+x_2[k]\\&\vdots\\x_{N-2}[k+1]&=b_{N-2}x[k]+a_{N-2}y[k]+x_{N-1}[k]\\x_{N-1}[k+1]&=b_{N-1}x[k]+a_{N-1}y[k]\end{aligned} \tag{1-75}$$

代入 $y[k]=x_1[k]+bx_0[k]$可得：

$$\begin{aligned}x_1[k+1]&=(b_1+b_0a_1)x[k]+a_1x_1[k]+x_2[k]\\&\vdots\\x_{N-2}[k+1]&=(b_{N-2}+b_0a_{N-2})x[k]+a_{N-2}x_{N-2}[k]+x_{N-1}[k]\\x_{N-1}(k+1]&=(b_{N-1}+b_0a_{N-1})x[k]+a_{N-1}x_{N-1}[k]\end{aligned} \tag{1-76}$$

此时的状态参数为：

$$\boldsymbol{A}\triangleq\begin{pmatrix}a_1&1&0&\cdots&0\\a_2&0&1&\cdots&\vdots\\\vdots&&\ddots&\ddots&0\\&\cdots&&0&1\\a_{N-1}&0&&\cdots&0\end{pmatrix},\quad \boldsymbol{b}\triangleq\begin{pmatrix}b_1+b_0a_1\\\vdots\\b_{N-1}+b_0a_{N-1}\end{pmatrix} \tag{1-77}$$

$$\boldsymbol{c}\triangleq[1,0,\cdots,0]^{\mathrm{T}},\quad d=b_0 \tag{1-78}$$

注意：在对偶现实中，状态传递矩阵 $\boldsymbol{A}$ 是式(1-70)的逆矩阵，$\boldsymbol{b}$ 和 $\boldsymbol{c}$ 进行了互换，d 的取值仍然是标量 b_0。

1.3　基于 MATLAB 的统计信号处理

尽管本书既讲到了连续时间随机过程，也讲到了离散时间随机序列，但所有基于 MATLAB[⊖] 的仿真试验都是关于离散时间序列的。在本节，我们简要介绍产生随机序列的一些方法，同时还介绍一些滤波函数，利用这些函数可以改变随机序列的性质和频域特性。

1.3.1　产生随机数的方法

MATLAB 中的随机数发生器可以用来产生伪随机序列 $X[k]$，这些样本在时间上是独立的，因此也是不相关的。有两种基本的随机数发生器：

⊖ MATLAB 是 Mathworks 公司(3 Apple Hill Drive，Natick，MA)的注册商标。

1）randn（N,1），产生一个$N\times1$的实随机向量，该随机向量服从零均值，方差为1的高斯分布。

2）rand（N,1），产生$N\times1$的随机序列，该序列在[0,1]区间上服从均匀分布。

图1-31给出了MATLAB中两种随机数发生器2000个样本的统计直方图，n表示落在x轴上某个特定区间的样本数量。如果将区间宽度减小，同时增加样本数量，那么直方图就会逐渐趋向于概率密度函数，

$$f_X(x)=\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\text{（标准高斯分布）} \tag{1-79}$$

$$f_X(x)=I_{[0,1]}(x)\text{（在单位区间上均匀分布）} \tag{1-80}$$

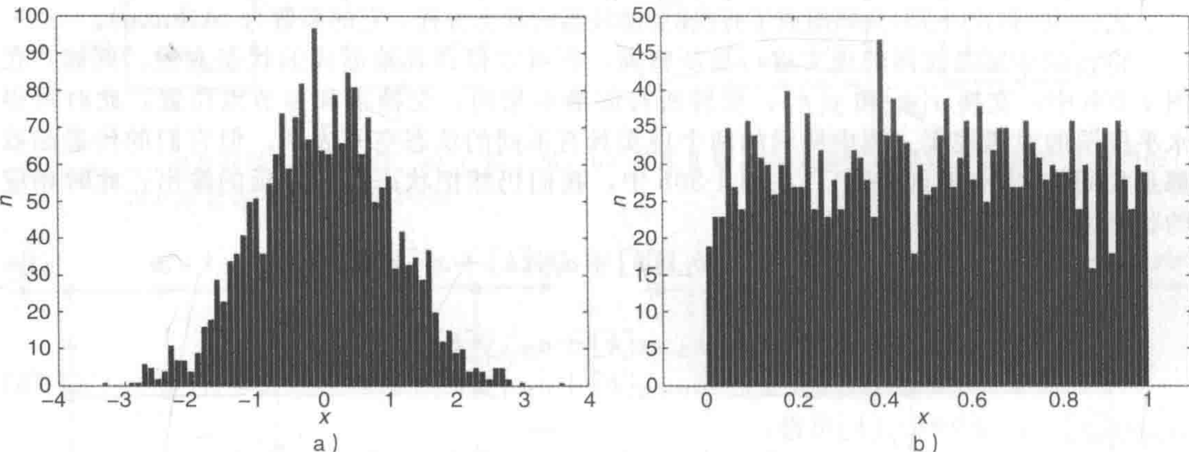


图1-31　MATLAB中两种随机数发生器样本的统计直方图（2000个样本）。a）标准正态分布；b）[0,1]区间的均匀分布

在式（1-80）中$I(x)$为指标函数（见附录B）。即使样本数目无限多，直方图和随机数发生器也只是一种近似，因为在计算机中使用的算法是有限精度的，不可能产生所有的实数。

对于一个高斯随机数发生器，如果想改变产生的随机数的均值和方差，那么可以按照下面的公式来修改随机序列：

$$Y[k]=\sigma_Y X[k]+\mu_Y \tag{1-81}$$

此时$Y[k]$仍然服从高斯分布，均值和方差分别为μ_Y和σ_Y^2。请注意在这个公式中$X[k]$首先是与σ_Y相乘，然后再加μ_Y，这与$\sigma_Y(X[k]+\mu_Y)$得到的结果是不一样的。利用函数rand和randn来产生的随机序列如图1-32所示。很明显，高斯现实的变化幅度大于均匀现实的变化幅度，均匀分布的取值严格限制在区间[0,1]中。不同时刻的样本相互之间是独立的，第5章将会介绍如何利用MATLAB产生相关序列。

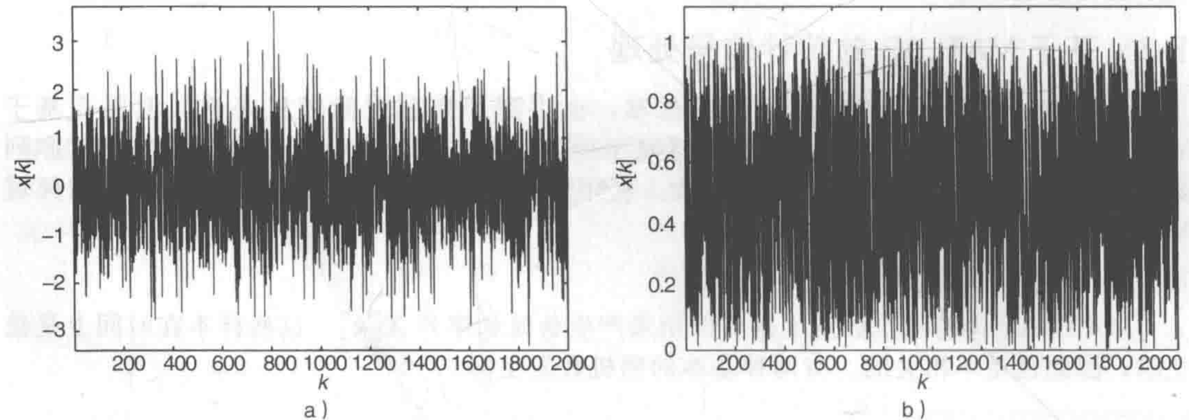


图1-32　非相关随机序列的现实。a）标准正态分布；b）[0,1]区间的均匀分布

随机数发生器 randn 非常有用，因为在实际情况中高斯过程是很常见的，比如通信系统中的噪声。将一个随机变量转变为另外一个随机变量是比较容易的，尤其是对于 rand 随机数发生器。如果 X 是一个 $[0,1]$ 区间的均匀分布随机变量，$F_Y(y)$是目标随机变量 Y 的累积分布函数，那么采用第 4 章介绍的变换方法，就可以轻松地通过如下的变换得到 Y：

$$Y = F_Y^{-1}(X) \tag{1-82}$$

其中 $F_Y^{-1}(y)$是 $F_Y(y)$的逆函数。高斯随机变量的累积分布函数及其逆函数如图 1-33 所示。在计算机仿真中，如果通过 rand 产生了 X 的一个样本，那么利用式(1-82)的变换就可以得到 Y。请注意在图 1-33 中，逆累积分布函数中 X 的变化范围为 rand 函数的支撑集 $[0,1]$。这并不是一个巧合，在第 3 章讨论过，Y 的累积分布函数为：

$$F_Y(y) \triangleq P(Y \leqslant y) \tag{1-83}$$

根据概率公理可知 $F_Y(y) \in [0,1]$（这点可从图 1-33看出）。$F_Y(y)$的支撑集是$\mathcal{R}$，尽管有的随机变量的支撑集是有限集，所以 $F_Y(y)$的取值为 0 或 1(对于均匀分布随机变量)。从图 1-33 的曲线可以看出，逆累积分布函数的值域是整个实线。rand 函数生成 $[0,1]$ 区间上的均匀分布，正好是累积分布函数的值域。通过逆累积分布函数对 rand 函数产生的样本进行映射，得到$\mathcal{R}$上相应的样本，也就是该逆累积分布函数定义的随机变量。在 MATLAB 中有很多种类型的随机数发生器，这些不同类型的随机数都可以通过调用 random("type", "parameters")来产生，其中"type"表示随机数的类型，"parameters"表示概率密度函数的参数(比如位置、尺度、形态、自由度、门限等)。

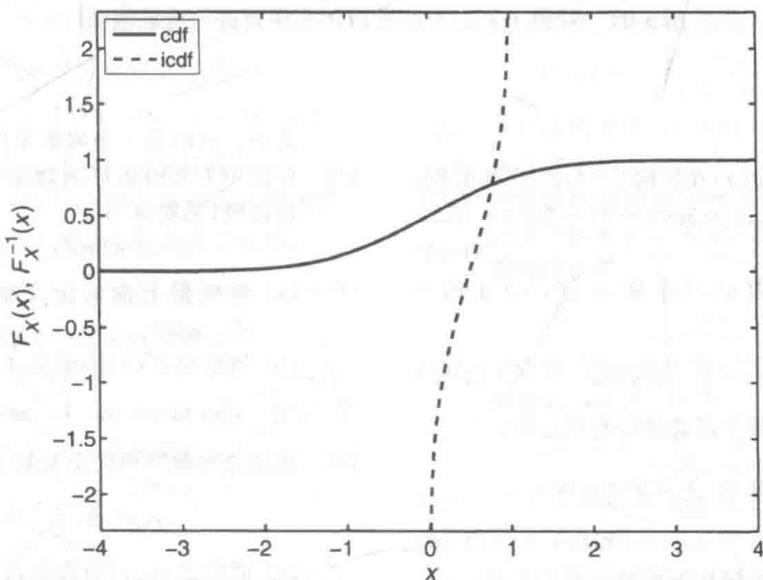


图 1-33　高斯随机变量的累积分布函数及其逆函数

1.3.2　滤波

MATLAB 中有两个基本函数可以用来对序列进行滤波：

- filer($\boldsymbol{b},\boldsymbol{a},\boldsymbol{x}$)：产生一个与输入向量 $\boldsymbol{x}$ 相同维数的输出向量，其中 $\boldsymbol{a} = [a_0,\cdots,a_{N-1}]^{\mathrm{T}}$ 和 $\boldsymbol{b} = [b_0,\cdots,b_{M-1}]^{\mathrm{T}}$ 表示式(1-66)中线性差分方程的滤波器系数。图 1-30a 中采用的是直接Ⅱ型。
- conv($\boldsymbol{h},\boldsymbol{x}$)：实现式(1-47)的卷积，产生一个长度为 $2N-1$ 的输出向量。其中求和限为 $\{0,\cdots,N-1\}$，有限长度的滤波器系数向量为 $\boldsymbol{h} = [h[0],\cdots,h[N-1]]^{\mathrm{T}}$，输入向量为 $\boldsymbol{x} = [x[k],\cdots,x[k-N+1]]^{\mathrm{T}}$。

如果令 $\boldsymbol{a}=1$ 同样也可以实现一个 FIR 滤波器。函数 conv 也可以实现一个 FIR 滤波器，不过与函数 filter 有所区别。因为 conv 输出向量的长度超过了输入向量 $\boldsymbol{x}$ 的长度，所以把 $\boldsymbol{x}$ 中那些"多出来"的 $x[k]$都假设为 0。因此，在许多滤波操作中都倾向于使用函数

filter。图 1-34 给出了函数 filter($\boldsymbol{b},\boldsymbol{a},\boldsymbol{x}$) 的 FIR 滤波器输出结果，输入为图 1-32a 中的高斯随机序列，其中 $\boldsymbol{b}=(1/10)[1,\cdots,1]^{\mathrm{T}}\in\mathcal{R}^{10},\boldsymbol{a}=1$。该滤波器具有低通的频率响应，而且我们可以看出输出没有输入显得那么"嘈杂"，该输出序列是相关序列。

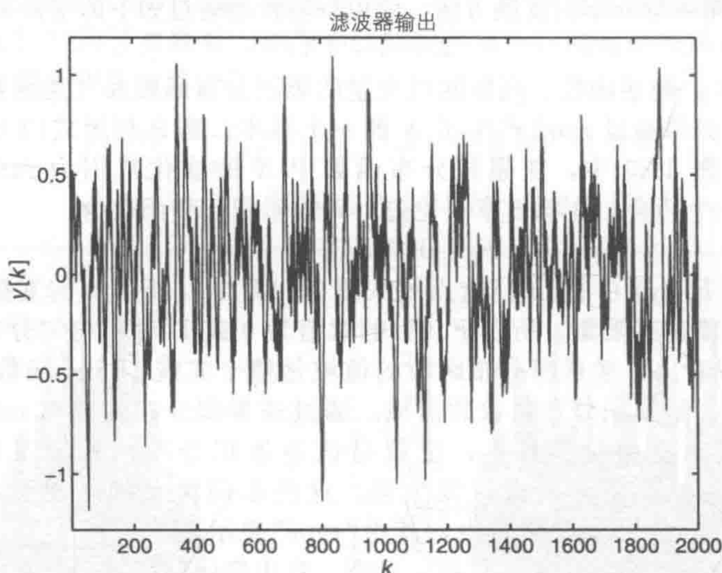


图 1-34 对图 1-32a)中的高斯随机序列进行滤波输出

习题

δ 函数

1-1 (a) 试求函数 $g(x)$在区间 $[-1,2.5]$ 的面积：

$$g(x)=5\delta(x)+3\delta(x-2)-2\delta(x-3) \tag{1-84}$$

(b) 试给出函数 $h(x)$在 $\mathbf{R}^{+}=[0,\infty)$ 上的一般表达式：

$$h(x)=\int_{-\infty}^{x}g(u)\mathrm{d}u \tag{1-85}$$

1-2 (a) 请对下面两个函数进行卷积运算：

$$g(x)=\sum_{n=1}^{3}2^{n}\delta(x-n),$$
$$h(x)=\exp(-x)u(x) \tag{1-86}$$

(b) 对下面两个函数再次进行卷积运算：

$$g(x)=\delta(ax+b),$$
$$h(x)=\exp(-x)u(x) \tag{1-87}$$

1-3 (a) 请用两个 δ 函数表示 $\delta(x^{2}-1)$。

(b) 请证明 $u(1/4-x^{2})$是一个矩形函数。

1-4 (a) 请对下面两个函数进行卷积运算：

$$g[k]=\sum_{n=1}^{3}2^{n}\delta[k-n],$$
$$h[k]=\sum_{n=1}^{2}\delta[k-n] \tag{1-88}$$

(b) 对下面两个函数再次进行卷积运算：

$$g[k]=\sum_{n=0}^{1}3^{n}\delta[k-n],$$
$$h[k]=(1/3)^{k}u[k] \tag{1-89}$$

其中，$u(k)$是一个离散单位阶跃函数。

1-5 可以用下面的矩形函数来对狄拉克 δ 函数进行建模(见附录 B)：

$$\delta(x)=\lim_{a\to 0}(1/a)\mathrm{rect}(x/a) \tag{1-90}$$

(a) 请根据上面的定义给出 $\delta(\alpha x)$ 的表达式，$\alpha\neq 0$。

(b) 请给出 $\delta'(x)$的定义，并证明：

$$\int_{-\infty}^{\infty}\delta'(x)\mathrm{d}x=0,\quad \int_{-\infty}^{\infty}x\delta'(x)\mathrm{d}x=-1 \tag{1-91}$$

1-6 狄拉克冲激序列的定义如下：

$$s(x)\triangleq\sum_{n=-\infty}^{\infty}\delta(x-n) \tag{1-92}$$

(a) 请给出 $s(\alpha x)$的表达式，$\alpha\neq 0$。

(b) 请证明下面的等式：

$$\delta(\sin(\pi x))=(1/\pi)s(x) \tag{1-93}$$

连续时间信号和系统

1-7 请计算下面系统的冲激响应 $h(t)$，其中输入为 $x(t)$，输出为 $y(t)$：

$$\frac{\mathrm{d}^{2}y(t)}{\mathrm{d}t^{2}}+2\frac{\mathrm{d}y(t)}{\mathrm{d}t}+y(t)=x(t) \tag{1-94}$$

1-8 已知系统的冲激响应如下，试推导该系统的单位阶跃响应：

$$h(t)=2\exp(-t)u(t)+\exp(-(t-5))u(t-5) \tag{1-95}$$

1-9 请计算下面信号的拉普拉斯变换，并给出它们的收敛域(ROC)，其中 erf(t)表示误差函数。

$$x_1(t)=\exp(-5|t|),\quad x_2(t)=\mathrm{erf}(t)u(t) \tag{1-96}$$

1-10　按照习题 1-9 的要求再次进行计算：

$$x_1(t)=\mathrm{rect}(t),\quad x_2(t)=1/t \tag{1-97}$$

1-11　按照习题 1-9 的要求再次进行计算：

$$x_1(t)=t\exp(-3t)u(t-1),$$
$$x_2(t)=\cos(2\pi t-5)u(t-2) \tag{1-98}$$

1-12　利用部分分式展开法计算拉普拉斯逆变换，其中收敛域为 $\mathrm{Re}(s)>-1$。

$$H(s)=\frac{2s}{s^2+3s+2} \tag{1-99}$$

1-13　按照习题 1-12 的要求再次进行计算，其中收敛域为 $-1<\mathrm{Re}(s)<2$。

$$H(s)=\frac{2s}{s^2-s-2} \tag{1-100}$$

1-14　对下面两个函数进行卷积：

$$x(t)=\exp(-2t)u(t)+\exp(-t+1)u(t-1),$$
$$h(t)=u(t-2) \tag{1-101}$$

1-15　计算下面系统的输出 $y(t)$，输入 $x(t)=\exp(-t)u(t-1)$。

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t}+4y(t)=x(t) \tag{1-102}$$

1-16　确定性协相关函数可由如下计算：

$$y(t)=\int_{-\infty}^{\infty}x(\tau)h(t+\tau)\mathrm{d}\tau \tag{1-103}$$

(a) 请利用 $X(\omega)$ 和 $H(\omega)$ 来计算 $Y(\omega)$。

(b) 请计算习题 1-14 中信号的协相关函数。

1-17　对于实值 $x(t)$，请证明帕斯塞尔定理。

$$\int_{-\infty}^{\infty}x^2(t)\mathrm{d}t=\int_{-\infty}^{\infty}|X(f)|^2\mathrm{d}f \tag{1-104}$$

1-18　请计算下面信号的傅里叶变换：

$$x_1(t)=5\mathrm{rect}(5t),$$
$$x_2(t)=\exp(-|t|)\cos(4\pi t) \tag{1-105}$$

1-19　请计算下面信号的频率响应 $Y(\omega)$：

$$y(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}\mathrm{d}\tau \tag{1-106}$$

离散时间信号和系统

1-20　计算下面系统的冲激响应 $h[k]$，其中输入为 $x[k]$，输出为 $y[k]$。

$$y[k]+2y[k-1]+4y[k-2]$$
$$=x[k]+2x[k-1] \tag{1-107}$$

1-21　已知系统的冲激响应如下，试推导该系统的单位阶跃响应：

$$h[k]=(1/3)^k u[k]-2u[k-1] \tag{1-108}$$

1-22　请计算下面信号的 z 变换，并给出它们的收敛域(ROC)：

$$x_1[k]=(k-1)^2u[k-1],$$
$$x_2[k]=(1/2)^k u[k]+4^k u[k+1] \tag{1-109}$$

1-23　按照习题 1-22 的要求再次进行计算：

$$x_1[k]=3^k u[-k+1],$$
$$x_2[k]=2u[k]-2u[k-5] \tag{1-110}$$

1-24　计算下面信号的离散时间傅里叶变换(DTFT)：

$$x_1[k]=\mathrm{sinc}[k-1],$$
$$x_2[k]=k(1/2)^{k-1}u[k] \tag{1-111}$$

1-25　对于下面给出的样本均值：

$$\overline{x}[k]=\frac{1}{k}\sum_{n=0}^{k}x[n] \tag{1-112}$$

(a) 当 $k=2$ 时，求出系统的零点。

(b) 试将 $\overline{x}[k]$ 写成 $\overline{x}[k-1]$ 的递归表达式，并用数字滤波器的结构来实现。

1-26　对下面两个函数进行卷积：

$$x[k]=a^{k-1}I_{\mathcal{N}}[k],$$
$$h[k]=u[k]-u[k-N] \tag{1-113}$$

其中 $I_{(a,b)}[k]$ 表示离散指示函数，$\mathcal{N}=\{1,2,\cdots\}$ 表示自然数。

1-27　确定性协相关函数可由如下计算：

$$y[k]=\sum_{n=-\infty}^{\infty}x[n]h[k+n] \tag{1-114}$$

请计算习题 1-26 中信号的协相关函数。

1-28　请证明 z 变换的导数特性，即

$$kx[k]\rightarrow -z\mathrm{d}X(z)/\mathrm{d}z$$

1-29　请用柯西留数定理(见附录 C)求下面 z 变换的逆变换，其中收敛域为 $|z|>1/2$。

$$H(z)=\frac{2(z+1)}{(z+1/2)(z-1/3)} \tag{1-115}$$

1-30　当收敛域为 $1/3<|z|<1/2$ 时，请再次求解习题 1-29。

1-31　请根据柯西留数定理来计算下面传递函数的逆变换 $h[k]$，要求 $h[k]$ 是一个左侧的冲激响应函数。

$$H(z)=\frac{3}{(z+1/2)(3/2-z)} \tag{1-116}$$

1-32　请根据 $H(z)$ 给出 DTFT 结果 $H(\mathrm{j}\omega)$ 的幅频和相频表达式：

$$H(z)=\frac{5z}{(z-1/4)(z+1/5)} \tag{1-117}$$

1-33　对于下面的有限长度序列，请再次求解习题 1-32。

$$x[k]=\exp(\mathrm{j}\omega_0 k)I_{\{0,\cdots,N-1\}}[k] \tag{1-118}$$

对连续时间信号进行采样

1-34　对连续时间信号 $x(t)=\exp(-\alpha t)u(t)$ 进行均匀采样，采样周期为 T_s，采样后的信号为 $x_s(t)=x(t)s(t)$，其中 $s(t)$ 表示理想冲激串(见式(1-43))。(a) 计算 $X_s(\omega)$ 的表达式，并证明 $X_s(\omega)$ 就是 $X(\omega)$ 以整数倍 $2\pi/T_s$ 重复出现的结果。(b) 推导采样信号 $x[kT_s]$ 的 DTFT 结果 $X(\mathrm{j}\omega)$，证明与 $X_s(\omega)$ 相比较，$X(\mathrm{j}\omega)$ 出现了混叠现象。

1-35　如果不采用理想冲激串，而改用下面的矩形脉冲串对 $x(t)$ 采样：

$$p(t)=\sum_{n=-\infty}^{\infty}\mathrm{rect}(t/T_p-nT_s) \tag{1-119}$$

其中 $T_p \ll T_s$。$p(t)$的傅里叶级数如下：

$$p(t) = \sum_{m=-\infty}^{\infty} p_m \exp(\mathrm{j}2\pi mt/T_s) \tag{1-120}$$

其中，

$$p_m = \frac{1}{T_s}\int_{-T_p}^{T_p} \exp(-\mathrm{j}2\pi mt/T_s)\mathrm{d}t \tag{1-121}$$

请给出 p_m 的闭合表达式，并以此给出采样波形 $x_p(t)=x(t)p(t)$的表达式及其傅里叶变换 $X_p(\omega)$。

1-36 对于下面的连续时间信号：

$$x(t) = 2\cos(200\pi t) + 4\cos(400\pi t) + 3\cos(600\pi t) \tag{1-122}$$

设用理想冲激串 $s(t)$对信号 $x(t)$进行采样，得到的采样波形为 $x_s(t)$。请计算 $x_s(t)$的傅里叶变换结果并绘制波形，采样率为(a) $f_s=300\text{Hz}$；(b) $f_s=500\text{Hz}$。请考虑在什么情况下会发生混叠。

1-37 理想重构滤波器的频率响应为：

$$H(\omega) = T_s \mathrm{rect}(\omega/2W) = \begin{cases} T_s, & |\omega| \leqslant W \\ 0, & \text{其他} \end{cases} \tag{1-123}$$

其中 W 是采样信号 $x(t)$的单边带宽。原始信号的傅里叶变换由 $X(\omega)=H(\omega)X_s(\omega)$可得，并假设满足奈奎斯特采样定理($f_s=1/T_s>2W$)。请用样本$\{x(kT_s)\}$和冲激响应$h(t)$相卷积的形式给出 $x(t)$的表达式。

离散时间滤波器

1-38 离散时间系统的传递函数为：

$$H(z) = \frac{5(z-2)}{(z-0.3)(z^2+0.1z-0.6)} \tag{1-124}$$

(a) 计算并绘制出直接Ⅰ型和Ⅱ型滤波器结构；

(b) 请分别使用部分分式展开法(PFE)和一阶向量乘积，计算并绘制出平行与级联滤波器结构。

1-39 对于下面的传递函数，按照习题1-38的要求进行计算。

$$H(z) = \frac{2(z^2-(5/6)z+1/6)}{(z+1/2)(z^2-1/16)} \tag{1-125}$$

1-40 离散时间滤波器的冲激响应为：

$$h[k] = (0.8)^k u[k] + (-0.4)^{k-1} u[k-1] \tag{1-126}$$

(a) 计算并绘制出该滤波器的直接Ⅱ型结构。

(b) 请以差分方程的形式，根据输入 $x[k]$写出输出 $y[k]$的表达式。

1-41 对于下面的系统：

$$H(z) = \frac{z+2}{(z^2-(1/12)z-1/12)(z+1/2)} \tag{1-127}$$

(a) 请以差分方程的形式，根据输入 $x[k]$写出输出 $y[k]$的表达式。

(b) 根据式(1-68)的定义，写出状态空间现实的参数$\{\boldsymbol{A},\boldsymbol{b},\boldsymbol{c},d\}$。

(c)根据式(1-75)的定义，重复计算(b)。

1-42 对于式(1-124)给出的系统，重复计算习题1-41。

仿真作业

1-43 利用 randn 函数产生1000个样本，并绘制出高斯序列 $x[k]$。请利用 filter 函数，根据下面的传递函数对 $x[k]$进行滤波操作产生输出序列 $y[k]$，并尝试修改参数 p。

$$H(z) = \frac{2}{z-p} \tag{1-128}$$

1-44 利用 rand 函数产生1000个样本，并绘制出均匀分布的序列 $x[k]$。请用 MATLAB 中的函数 icdf 将 $x[k]$转变为一个参数为 λ 的指数随机序列。对于不同取值的参数 λ，请比较输出序列的取值分布情况。(a) $\lambda=1$；(b) $\lambda=0.1$。

1-45 利用 randn 函数产生10个高斯随机变量，通过 sign 函数将其映射为±1，然后利用 upsample 函数在每个比特之间插入0，从而得到一个包含20个符号的0和±1的序列。用 rectpulse 函数将每个符号映射为包含100个样本的矩形函数，然后用 stairs 函数将这些脉冲幅度调整(PAM)波形绘制出来。

进一步阅读

本章最后部分介绍的信号和系统方面的知识，可以在以下文献中得到进一步探讨。

狄拉克 δ 函数：Bracewell(1978)，Weisstein(2003)。

连续时间线性系统：Lathi(1965)，Lathi(2000)，Oppenheim，Willsky 和 Nawab(1996)。

数字信号处理：Hamming(1983)，Oppenheim 和 Schafer(2009)，Mitra(2010)，Rabiner 和 Gold(1975)，Stearns 和 Hush(1990)。

状态空间的表示：Kailath(1980)，Mendel(1995)，Tretter(1976)。

基于 MATLAB 的信号处理：Chaparro(2011)，Childers(1997)，Ingle 和 Proakis(2011)，McClellan，Schafer 和 Yoder(1998)。

第一部分

概率、随机变量与期望

第2章

概　率　论

2.1　引言

本章介绍基于三大公理的概率论，这与我们对事件概率的直观认识是相符的。首先需要了解集合理论方面的一些背景知识，包括事件和域的定义，这样就可以确定和计算概率。随机试验中最简单、最常见的一个例子可能就是硬币投掷试验。尽管这个试验非常简单，但我们可以用它来给出随机事件的定义和性质。在这个试验中只有正面(H)和反面(T)这两种可能的输出，因此这也可以用作数字通信系统的模型。在数字通信系统中经常采用二元字母表$\{0,1\}$(如果希望输出结果关于0对称，也可采用$\{-1,1\}$)。尽管硬币投掷试验看上去没有什么内涵，但在许多实际应用中却是非常有用的。

掷骰子试验也常常被用来研究概率的各种特性，因为这个试验具有多种可能的输出结果$\{1,2,3,4,5,6\}$，因此它比硬币投掷试验要复杂一些。尽管任何一个多输出值的试验都是可以的，但是由于大多数读者对掷骰子试验都很熟悉，所以我们选择这个试验。这个例子与通信也是有关联的，比如可以作为脉冲幅度调制(PAM)的模型。在PAM模型中，通过信道传输的符号是从一个有限数目字母表中选取出来的，比如4，8，16等。

根据上面介绍的两个简单试验，我们可以很好的描述构建概率空间的一些重要原则，在这个概率空间中试验的输出结果是有限多的。通过深刻理解这些例子，概率模型可以被扩展到具有无限输出结果的试验。根据输出结果的数目，可以将随机试验划分为三类：(ⅰ)输出结果的数量有限(比如一次硬币投掷试验)；(ⅱ)输出结果的数量无限但是可数(比如无限进行下去的硬币投掷试验)；(ⅲ)输出结果为连续值(不可数的，比如对温度的测量)。这三种基本类型的输出结果可以组成更多更复杂的随机试验。

很显然，也应该分类计算随机试验中事件的概率。(ⅰ)将结论直接拓展到该类试验；(ⅱ)需要进行一些适当的运算，由于输出结果是可数的，可以建立一个输出结果与自然数集合$\mathcal{N}=\{1,2,\cdots\}$之间的一一映射。(ⅲ)计算这类试验的概率是最复杂的，因为输出结果是连续的。任取两个值$a<b$，总存在一个$\varepsilon>0$使得$a<a+\varepsilon<b$。此时，对于输出结果为无限不可数的情况，该如何计算概率值也不甚明了。实际上，连续试验中某一特定的输出结果的概率为0。对于连续试验中的事件，需要专门考虑该如何定义它的概率空间。

在此我们总结一下该如何计算上面三种随机试验的概率：

- *有限*。将感兴趣事件中输出结果的数目比上样本空间中输出结果的总数目，比值就是该事件的概率。此处需要用到组合数学中的一些基本规律。
- *无限可数*。对离散概率质量函数进行*求和*来计算概率，离散概率质量函数可以很好地反映这种随机试验的特性。计算无限可数的问题，需要用到有限求和以及*无限求和*的一些方法。
- *无限不可数*。对连续概率密度函数进行*积分*来计算概率，连续概率密度函数可以很好地反映这种随机试验的特性。计算无限不可数的问题，需要用到*微积分*的一些方法。

解决无限可数问题的那些方法，当然也可以用到有限输出的情况，不过为了简便一般都用简单求和的方法。无限可数和无限不可数的问题处理起来比较麻烦，需要用到一些更加精确的特性，比如由概率质量函数(pmf)和概率密度函数(pdf)提供的特性。随后将会学习用微积分的方法来处理无限可数问题(利用狄拉克δ函数)，但其本质是有限(无限)求和问题。

表 2-1 给出了三种随机试验的一些例子。很显然，投掷两次硬币所有可能的输出结果为：$\{HH, TT, HT, TH\}$。如果我们想计算 H 至少出现过一次的概率，那么只需要将这种输出的次数(3 次)除以总的输出次数。可知。H 至少出现过一次的概率为 0.75，计算公式为 P(H 至少出现一次)$=P$(HH 或 HT 或 TH)$=0.75$。对于表 2-1 中无限可数的例子，我们希望计算在第三次投掷时才出现 H 的概率。这个例子显然比前面有限输出的例子复杂一些，我们需要考虑所有类似于 $TTHT\cdots$，$TTHH\cdots$，$TTTH\cdots$这样的序列。后面将会学习用概率质量函数来直接计算这个概率，这显然比试图统计所有可能的结果简单。同样地，对于无限不可数的问题(连续试验)，可以用概率密度函数计算概率，即对感兴趣的输出结果进行积分，而这些事件定义在实数区间$\mathcal{R}$上的一段区间内。

表 2-1 基本随机试验的一些例子

输出	有限	无限
离散	两次投掷硬币试验(有限)	重复硬币投掷试验(无限可数)
连续	无	温度测量(不可数)

尽管用“试验”这个词语来描述随机结果是如何产生的，但在实际情况下“试验”这个用语有点人造的意思，因为这些事件都是“刚刚发生的”(比如温度的起伏、放射性衰变等)。但是，我们用“试验”这个词语并不只是针对人造试验(比如投掷骰子，或者计算机编程产生随机数)，也可用于自然事件。对于后面这种情况，可以把试验看做是底层机理的一种模型产生的事件。

在实际的物理世界中，用概率模型来描述一个事件是非常有用的，因为绝大部分现象用物理模型来准确描述都是非常复杂的。比如在第一章中提到的单次硬币投掷试验，如果要对其进行建模就需要考虑很多因素，比如投掷的高度、速度和角度、硬币着陆处的材质等等。很显然，用这样的模型来对这个试验建模非常复杂，而且毫无疑问也是不够的。如果我们采用概率模型，就不需要准确的物理模型，而且还可以对事件的发生作出相对准确的预测，这与我们对随机特性的直观认识是一致的(在第 1 章中提到的某输出结果的发生频率)。概率模型是非常有用的，我们可以根据它来“谈谈”某个即将发生的事件的一些性质，尽管这个事件是由诸多非常复杂的底层物理因素共同决定的。

2.2 集合与样本空间

首先给出集合与样本空间的定义，大部分的读者都需要再复习一下这方面的知识。

定义(集合(Set)) 对象或者代表对象的数字的全体称作**集合**。这些对象称作集合的元素或点。

在许多情况下，可以用遍历的方法来描述一个集合中的所有对象或试验输出，但是对于大集合这种方法常常是不可取的。实际上，我们一般使用数字来代表集合中的所有元素，比如整数、实数或复数，这样在数学上处理起来比较方便，尤其是在第 3 章中定义随机变量时。例如对于单次硬币投掷试验，用数字来表示输出两种结果将非常方便，即 $1\equiv H$ 和$0\equiv T$。

集合一般采用拉丁字母表前面那些符号的大写形式来表示，请参见下面的这些例子：

例 2-1 集合 $A=\{2,4,6,8\}$ 包含 4 个整数，集合 $B=\{\cdots,-1,0,1,2,\cdots\}$ 包含所有整数。◀

一般用$\mathcal{Z}$来表示所有整数组成的集合，$\mathcal{Z}^+$包含所有的正整数(包含 0)。

例 2-2 集合 $C=(1,5)$ 包含 1 和 5 之间的所有实数(不包含 1 和 5)，集合 $D=[1,5]$ 包含 1 和 5 之间的所有实数(包含 1 和 5)。◀

一般用$\mathcal{R}$来表示所有实数组成的集合，$\mathcal{R}^+$包含所有的正实数(包含 0)。我们也经常使

用$\pm\infty$这个符号，但实际上它们并不是实数。除了在本书的附录部分，复数的情况并不经常遇到，我们一般用$\mathcal{C}$来表示所有复数组成的集合。

$a\in A$ 表示元素 a 属于集合 A；元素 b 不属于集合 A 写为 $b\notin A$。可以用集合中元素的数量来对集合进行分类。

定义(集合的基数(Cardinality)) **集合的基数**表示集合中元素的数目。用 $|E|$ 表示集合 E 的基数。

一个可数集元素的数目可以是有限的，也可以是无限的，前者称作有限集(集合的基数$<|\mathcal{N}|$)，后者称作无限可数集(集合的基数$=|\mathcal{N}|$)。无限可数集中的元素与自然数集合 $\mathcal{N}\triangleq\{1,2,\cdots\}$ 是一一对应的(请注意，自然数集合$\mathcal{N}$不包括 0)。

例 2-3 $\mathcal{Z}$和$\mathcal{Z}^+$与自然数集$\mathcal{N}$有相同的基数，因此它们都是无限可数的。具有 a/b 形式的数都是有理数，其中 a，$b\in\mathcal{Z}(b\neq0)$，有理数集$\mathcal{Q}$的基数也是$|\mathcal{N}|$。◀

例 2-4 无理数集是不可数的，其基数大于$|\mathcal{N}|$。类似的，实数集$\mathcal{R}$以及定义在$\mathcal{R}$上的任意一段区间都是不可数的，基数都大于$|\mathcal{N}|$。“几乎所有的”无理数都是超越数，π 和 e 是最著名的无理数。◀

在例 2-1 中，A 是有限集，集合 B 是无限不可数的。如果一个无限集合是不可数的，那么它就是一个不可数集合且基数大于$|\mathcal{N}|$。在例 2-2 中集合 C 就是不可数的。可数集合也被称作是离散集合，不可数集合称作连续集合。

集合的这些区别在后面推导随机试验事件的概率空间中用处很大。非常有必要准确定义随机试验的事件以及相应的代数运算，从而给出统一的概率测度。图 2-1 用简单的例子很好的总结了三种基本类型的集合。同时也给出了一个混合集合的例子，这个集合同时包含了离散和连续的分量。根据定义可知混合集合是不可数的(无论它是否包含离散分量，只要集合中包含了连续分量就是不可数的)，但在第 3 章讨论随机变量的时候会发现：根据是否包含有限分量或无限不可数的分量，可以非常方便地对集合进行分类。图 2-1 中，运算符$\cup$表示取并集，稍后将学习集合的有关操作。

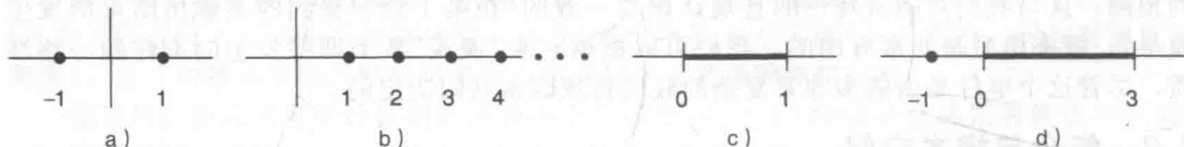


图 2-1 三种基本类型的集合，以及混合集合。a) 有限 $A=\{-1,1\}$；b) 无限可数 $B=\{1,2,3,4,\cdots\}$；c)无限不可数(一段区间) $C=[0,1]$；d)混合型(一段区间和一个点) $D=\{-1\}\cup[0,3]$

我们一般采用大括号$\{\cdot\}$来表示离散集合(无论是有限集或者无限可数集)，比如例 2-1。但是从下面的例题可知，用函数来表示连续集合的时候采用大括号也是非常方便的。

例 2-5 离散集：$A=\{x^3-1:x=0,2,4\}=\{-1,7,63\}$。连续集：$B=\{x:x\geqslant-1\}=[-1,\infty)$。新定义的集合 $C=\{x:x\in A \text{ 或 } x\in B\}=A\cup B$。◀

上文的冒号(:)读作“使得”(such that，缩写为 *s.t.*)。最后一个例子描述了取并集的操作(或者$\cup$)：C 是所有 x 组成的集合，在这里 x 属于集合 A，或者属于集合 B。可以用圆括号$(\cdot)$和方括号$[\cdot]$来描述不可数集合，如例 2-2 和 2-5 所示。接下来将要定义不可数集合中各种类型的区间。

定义(区间与单元素集合) 开区间 (a,b) 表示 a 和 b 之间所有实数组成的集合($a<b$)，但不包括边界点 a 和 b。闭区间 $[a,b]$ 表示 a 和 b 之间所有实数组成的集合($a<b$)，包括边界点 a 和 b。半开半闭区间 $[a,b)$ 与 $(a,b]$ 由闭区间 $[a,b]$ 得到，只不过要排除一个边界点：$[a,b)$ 不包括 b，$(a,b]$ 不包括 a。这些边界点称作**单元素集合**，它们可以是实数集合$\mathcal{R}$中的任意独立点。

我们也可以用下面的括号符号来定义区间(见例 2-5)：

$$(a,b)\triangleq\{x:a<x<b\},\quad (a,b]\triangleq(x:a<x\leqslant b\}$$
$$[a,b)\triangleq\{x:a\leqslant x<b\},\quad [a,b]\triangleq\{x:a\leqslant x\leqslant b\} \tag{2-1}$$

因为区间是一个连续集合，因此可以用数学方法来定义开区间。如果 $x\in(a,b)$，那么一定存在一个 $\varepsilon>0$，使得 $x+\varepsilon\in(a,b)$ 以及 $x-\varepsilon\in(a,b)$。无论 x 离 a 或 b 有多近，总存在这么一个 ε 在 $x\in(a,b)$ 与点 a 和 b 之间。当然，这个性质在半开半闭区间或闭区间中并不成立，因为可以把边界点选作 x。

为了能够详细描述事件及其性质，我们在此定义试验中的基本要素。

定义(样本空间 Ω) 随机试验中所有可能的输出组成的集合为**样本空间** Ω。样本空间也称作底集(universe)或全集(universal set)。用 ω 来表示 Ω 中的一个元素，该元素也可以用下标来进行区分，如 ω_n。

例 2-6 $\Omega=\{0,1\}$ 可以用来表示二元数字通信系统可能的符号，也可以用来表示单次硬币投掷试验($H\equiv1$，$T\equiv0$)组成的样本空间。◀

例 2-7 $\Omega=[0,\infty)=\mathcal{R}^+$ 表示一个试验中无数个非负实数输出组成的集合。$\Omega=(-\infty,\infty)$ 表示所有实数$\mathcal{R}$组成的集合。请注意，我们在这里总是用开区间(不包括$\pm\infty$)，因为$\pm\infty$不是实数。◀

对于全集 Ω 中的一系列集合，可以对它们之间的关系进行定义。

定义(子集 $A\subset B$) 如果 A 是 B 的**子集**，那么 A 中的所有元素都属于 A，然而 B 中的元素不一定属于 A。如果 $A\subseteq B$，那么 A 和 B 有可能是相同的集合。

例 2-8 如果 $A=[0,5]$，$B=[0,10)$，那么 $A\subset B$。半开半闭区间 $(a,b]$ 是闭区间 $[a,b]$ 的子集。很显然，任意一个集合都是自身的子集：$A\subset A$。如果 B 中所有的元素都属于 A，这两个集合是相等的。◀

定义(集合相等 $A=B$) 如果 A 集合和 B 集合的元素相同，即 $A\subset B$ 同时 $B\subset A$，那么 A 和 B 是**相等**的。

因为样本空间 Ω 包含试验中所有可能的输出结果，因此我们可以描述一个特定集合之外的元素。

定义(补集 A^c) 全集 Ω 中所有不属于 A 的元素组成的集合就是 A 的**补集**。因此 $a\notin A$ 一定有 $a\in A^c$。补集可以标记为：$A^c=\{x:\ x\notin A\}$。

例 2-9 如果 $\Omega=[0,100]$，$A=[0,25)$，那么 $A^c=[25,100]$。如果 $\Omega=\mathcal{R}$，$B=(0,\infty)$，那么 $B^c=(-\infty,0]$。◀

A 的补集也可以写成 $\overline{A}$，在这里我们采用 A^c 的标记，因为后面会用 $\overline{A}$ 表示样本矩(比如样本均值)。请注意必须首先定义 Ω，这样 A^c 才有意义。最后介绍两个重要的集合。

定义(空集 ϕ) **空集**是指没有任何元素的集合。空集有时也称作零集(null set)。根据前面的定义可知 $\phi=\Omega^c$。

有的时候也用{}来表示空集。空集的意思是“没有任何事件发生”。从后面的介绍可知，空集的概念在定义集合的代数运算时将要用到。

定义(幂集$\mathcal{P}(\Omega)$) **幂集**包含样本空间 Ω 中所有的子集，同时也包含空集 ϕ 和样本空间 Ω。同时还可以定义任意子集 $E\subset\Omega$ 的幂集，标记为$\mathcal{P}(E)$。

例 2-10 如果 $\Omega=\{2,4,6\}$，那么 $\mathcal{P}(\Omega)=\{\phi,\Omega,2,4,6,\{2,4\},\{2,6\},\{4,6\}\}$ 中包含 8 个元素。对于子集 $E=\{2,4\}\subset\Omega$，其幂集 $\mathcal{P}(E)=\{\phi,E,2,4\}$。对于一个无限可数集合(比如自然数集$\mathcal{N}$)，它的幂集是无限不可数的，幂集的基数与实数集$\mathcal{R}$的相同。◀

如果 E 是有限集，那么$|\mathcal{P}(E)|$也是有限的。

命题 2-1 如果集合 E 包含 N 个元素，那么 E 的幂集的基数为$|\mathcal{P}(E)|=2^N$。

证明：可根据附录 E 中的二项式定理证明这个命题，即

$$(x+y)^N = \sum_{n=0}^{N} \binom{N}{n} x^{N-n} y^n \tag{2-2}$$

其中

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{2-3}$$

式(2-3)是二项式的系数(在后续的有限样本空间的概率中还会讨论)。

二项式系数是指从 N 个元素中选取含 n 个元素的子集的数目(即组合)。对所有可能的 n 进行累加，并设式(2-2)中的 $x=y=1$ 可得$\mathcal{P}(E)$中元素的数目：

$$\sum_{n=0}^{N} \binom{N}{n} = (1+1)^N = 2^N \tag{2-4}$$

命题得证。

根据这些集合的基本定义，就可以定义集合的运算。

2.3 集合的运算

绝大多数概率问题都涉及对集合的运算。两种最基本的集合运算就是取并集和取交集。

定义(取并集 $A\cup B$) 两个集合的**并集**包含 A 或 B 中的所有元素。

$$A \cup B = (x{:}x \in A \text{ 或 } x \in B) \tag{2-5}$$

有的时候也用求和 $A+B$ 来表示取 A 和 B 的并集。

定义(取交集 $A\cap B$) 两个集合的**交集**的元素既属于 A 同时也属于 B。

$$A \cap B = \{x{:}x \in A \text{ 同时 } x \in B\} \tag{2-6}$$

有时也用乘积 AB 来表示取 A 和 B 的交集。

例 2-11 如果 $A=\{1,2,3\}$，$B=\{3,4\}$，那么 $A\cup B=\{1,2,3,4\}$，$A\cap B=\{3\}$。如果 $E=[2,\infty)$，$F=(-\infty,5]$，那么 $E\cup F=\mathcal{R}$，$E\cap F=[2,5]$。如果 $\Omega=\mathcal{R}$，$A=[a,b]$，$B=[a,b]$，那么 $AB=A$，$A\cup B=B$，$A^cB=\{a,b\}$ 为有限集，$AB^c=\phi$。◀

可以将取并集和取交集的运算拓展到任意数量的集合，可以包含无限多个集合。

$$B_1 = \bigcup_{n=1}^{N} A_n, \quad B_2 = \bigcap_{n=1}^{N} A_n \tag{2-7}$$

为方便起见，一般使用 AB 来表示取交集，但我们仍然用 $A\cup B$ 来表示取并集，因为 $A+B$ 可能会与事件概率相加相混淆。

很显然，取并集和取交集操作满足下面的基本定理：

- 交换律

$$A \cup B = B \cup A, \quad AB = BA \tag{2-8}$$

- 结合律

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (AB)C = A(BC) \tag{2-9}$$

- 分配律

$$A \cup (BC) = (A \cup B)(A \cup C), \quad A(B \cup C) = (AB) \cup (AC) \tag{2-10}$$

这些运算还满足德摩根定律，该定律将取交集、取并集和取补集运算综合在一起。

- 德摩根定律

$$(A \cup B)^c = A^cB^c, \quad (AB)^c = A^c \cup B^c \tag{2-11}$$

这些定理可以用维恩图法得到验证(维恩图稍后将会学习)，而且它们也可以拓展到任意数量的集合。

接下来，讲解对区间取并集和取交集运算的一些特性。在连续样本空间上，取无限并集和交集并不直观。对于合适取值的$\{a_n\}$，$\{b_n\}$，c 和 d，对实线空间$\mathcal{R}$上的区间取并集和取交集有如下特性：

- 对闭区间取有限并集或交集得到闭区间：

$$\bigcup_{n=1}^{N}[a_n,b_n]=[c,d],\quad \bigcap_{n=1}^{N}[a_n,b_n]=[c,d] \tag{2-12}$$

- 对开区间取有限并集或交集得到开区间：

$$\bigcup_{n=1}^{N}(a_n,b_n)=(c,d),\quad \bigcap_{n=1}^{N}(a_n,b_n)=(c,d) \tag{2-13}$$

- 对开区间取无限可数的并集得到开区间：

$$\bigcup_{n=1}^{\infty}(a_n,b_n)=(c,d) \tag{2-14}$$

- 对闭区间取无限可数的交集得到闭区间：

$$\bigcap_{n=1}^{\infty}[a_n,b_n]=[c,d] \tag{2-15}$$

前面两个特性是关于有限数目的开区间或闭区间，因此很容易得到验证。后面两个特性也可以直接得到验证，尤其是理解了我们接下来要讨论的这两个不太直观的性质。上面介绍的四种运算得到的区间与原始区间的特性是相同的。

- 对闭区间取无限可数的并集可得到任意性质的区间(闭、开、半开半闭)：

$$\bigcup_{n=1}^{\infty}[a_n,b_n]=(c,d)\text{或}[c,d]\text{或}[c,d)\text{或}[c,d] \tag{2-16}$$

- 对开区间取无限可数的交集可得到任意性质的区间(闭、开、半开半闭)：

$$\bigcap_{n=1}^{\infty}(a_n,b_n)=[c,d]\text{或}(c,d]\text{或}[c,d)\text{或}(c,d) \tag{2-17}$$

上面介绍的这两个性质非常有趣，因为通过集合运算后得到的区间性质竟然与运算之前的区间性质不一样。因此，无法对闭区间取无限并集，或者对开区间取无限交集给出任何一般性的结论。下面的两个例题将进一步描述这个性质。

例 2-12 如果 $(a,b_n)=(-1,1/n)$，那么 $\bigcap_{n=1}^{\infty}(a,b_n)=(-1,0]$。可以看出所有的开区间都包含0，它们的交集当然也包含0(即使是对无限多个区间取交集)，最终的结果得到的是一个半开半闭区间。 ◀

例 2-13 如果 $[a_n,b]=[1/n,1]$，那么 $\bigcap_{n=1}^{\infty}[a_n,b]=(0,1]$。很显然，每个区间都不包含0，因此它们的并集也不会包含0，即使是无限多个区间取并集。最终得到的是一个半开半闭区间。 ◀

还可以用很多例题来描述式(2-16)和式(2-17)中其他三种类型的区间。为方便起见，表格2-2总结了区间的集合运算所有可能的结果。表格的输入端是各种类型的区间，这表明任意一种类型的区间都可以参与这些集合运算(尽管不一定有必要)，但结果只取决于第一列的这些原始区间是如何定义的。我们在此归纳这些结果是因为集合代数必须是闭合的，包括取并集和取交集运算，而且 σ 域(稍后会在实数空间 $\mathcal{R}$ 定义)不能仅仅包含一种类型的区间。

表 2-2 区间的集合运算

区间类型	有限并集	有限交集	无限可数并集	无限可数交集
$[a_n,b_n]$	$[c,d]$	$[c,d]$	$(c,d),(c,d],[c,d),[c,d]$	$[c,d]$
$[a_n,b_n)$	$[c,d)$	$[c,d)$	$(c,d),[c,d)$	$[c,d),[c,d]$
$(a_n,b_n]$	$(c,d]$	$(c,d]$	$(c,d),(c,d]$	$(c,d],[c,d]$
(a_n,b_n)	(c,d)	(c,d)	(c,d)	$(c,d),(c,d],[c,d),[c,d]$

取交集运算在定义下面的这个集合性质时非常有用。

定义(互斥性(mutually exclusive)) 如果集合 A 和 B 没有共同的元素，即 $AB=\phi$，那么 A 和 B **互斥**。集合互斥也称作**不相交**。

上面的这个定义可以拓展到任意数量的集合$\{A_n\}$，只要满足任意两个集合不相交即可(称作两两不相交)：$A_nA_m=\phi$，$m\neq n$。此外，一个样本空间可以被划分为若干个不相交的集合，这些集合中的所有元素组成样本空间 Ω。这些集合称作样本空间的划分(partition)。

定义(划分(partition)) 如果一系列不相交的集合$\{A_n\}$满足$\bigcup_n A_n=\Omega$，那么$\{A_n\}$就称作样本空间 Ω 的**划分**。

一个样本空间划分并不是唯一的。当且仅当样本空间 Ω 是有限的，其划分才是有限的。如果集合$\{A_n\}$是相关的，但仍满足$\bigcup_n A_n=\Omega$，那么$\{A_n\}$被称作完全穷尽的(collectively exhaustive)。

用维恩图可以非常方便地描述集合间的各种关系。维恩图是集合的一种非常简单的图形化表示方法，一般用圆圈来表示样本空间内的集合，用这些圆圈的重合程度来表示集合间的相互关系。图 2-2a 表示的是 $A\cap B$ 的关系。请注意：如果 A 和 B 是不相交的，那么 $A\cap B=\phi$，同时在这个例子中，A 和 B 之外没有任何元素，因此 $A\cup B=\Omega$。在图 2-2b演示的是将样本空间划分为三个不相交的集合$\{A_1,A_2,A_3\}$。这些圆圈没有任何相互重叠的部分，样本空间中的所有元素也都包含在这三个集合里，因此 $A_1\cup A_2\cup A_3=\Omega$。

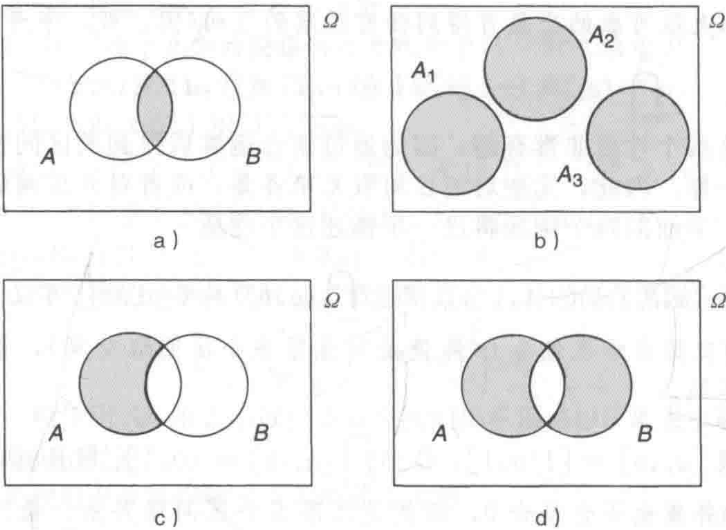


图 2-2 利用维恩图来直观的解释集合运算(阴影区域)。a) $A\cup B$；b) 划分：$A_1\cup A_2\cup A_3=\Omega$；c) 差：$A-B$；d) 排斥：$A\oplus B$

在本节最后，我们再给出两个非常有用的集合运算的定义。

定义(差 $A-B$) 集合 A 和 B 的差是指去除 A 与 B 的共同元素后，A 中剩余元素组成的集合。也可以用前面定义的集合运算表示为 $A-B=AB^c$。有的时候也用 $A\setminus B$ 来表示两个**集合的差**。

图 2-2c 用维恩图来表示了两个集合的差。如果 A 和 B 是不相交的，那么 $A-B=A$，$B-A=B$。差运算不满足交换律，即 $A-B\neq B-A$(除非 $A=B$，此时 $A-B=\phi$)。两个集合的差并不是一个对称运算。用维恩图也可以很方便的验证 $A-B=AB^c$。

定义(排斥 $A\oplus B$) 除了不包含 A 和 B 的共同元素，排斥运算得到的集合包含 A 和 B 的所有元素：$A\oplus B=A\cup B-AB=(A-B)\cup(B-A)$。根据最后这个公式，排斥运算也称作**对称差**(symmetric difference)。

图 2-2d 给出了如何用维恩图来表示 $A\oplus B$ 的例子。

2.4 事件与域

样本空间 Ω 包含一个随机试验所有可能的输出结果。这些输出结果称作 Ω 的元素(或点)。但是在很多情况下，我们并不关心某一个输出结果，只关心 Ω 的一个子集，即事件(events)。

定义(事件) 一个随机试验的**事件** E，是指样本空间 Ω 中具有某种共同特性的元素组成的子集。

例 2-14 将两枚硬币同时投掷出去存在四种可能的输出结果，这四个元素组成的样本空间 Ω 如图 2-3 所示。考虑至少存在一个 H 的事件，即 $E=\{HH,HT,TH\}$。可以看出，E 并不是一个特定的输出，它只是三个有共同特征的可能的输出结果。很显然，E 是 Ω 的一个子集。 ◀

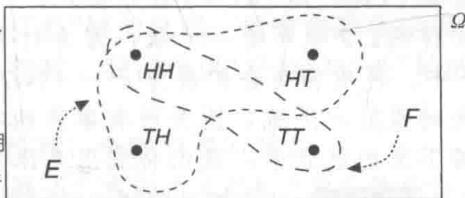


图 2-3 同时投掷两枚硬币，在样本空间中的两种基本事件

例 2-15 单次投掷骰子试验得到的样本空间为 $\Omega=\{1,2,3,4,5,6\}$，那么 $E=\{1,3,5\}$ 表示结果为骰子点数为奇数的事件。 ◀

例 2-16 在一次试验中同时投掷两个骰子，两个骰子点数相同的事件为子集 $A=\{\{1,1\},\{2,2\},\{3,3\},\{4,4\},\{5,5\},\{6,6\}\}$。 ◀

例 2-17 如果 $\Omega=[0,1]$，那么 $A=[0,1/2]$，以及 $B=[0,1/4]\cup[5/8,7/8]$ 为 Ω 的两个事件。元素 $\omega_n\in\Omega$ 对应于一个输出结果，这也是一种特定类型的事件。 ◀

定义(基本事件(Elementary Event)) Ω 中只包含一个元素的子集称作**基本事件**，基本事件等同于一个输出。

有的时候，基本事件也被称作原子事件(atomic event)，但我们将“原子”这个名字保留给后面将要介绍的 σ 域的特定要素。

例 2-18 例 2-15 中的子集 $F=\{1\}$，以及例 2-16 中的子集 $\{2,2\}$ 都是基本事件。 ◀

我们的目的是为了计算随机试验的事件的概率，因此需要合理定义事件使之满足各种代数定律，并且与样本概率空间保持一致。为了使事件的运算严密，我们在此给出域(*field*)的定义，即满足某种代数条件的事件簇。根据这个定义，我们就可以把 Ω 中特定的子集包括进来，这对于连续样本空间 Ω 尤为重要。

定义(域 $\mathcal{F}$) **域**是指一个试验中事件的非空集合，并且满足以下特性：

(ⅰ) 如果 $A\in\mathcal{F}$，那么 $A^c\in\mathcal{F}$。

(ⅱ) 如果 $A_n\in\mathcal{F}$，$n=1,2,\cdots,N$，N 是有限值，那么 $\bigcup_{n=1}^{N} A_n\in\mathcal{F}$。

域也称作事件代数(an algebra of events)。从上面的条件可知以下性质成立：

(ⅲ) $\Omega\in\mathcal{F}$。

(ⅳ) $\phi\in\mathcal{F}$。

(ⅴ) 如果 $A_n\in\mathcal{F}$，$n=1,2,\cdots,N$，N 是有限值，那么 $\bigcap_{n=1}^{N} A_n\in\mathcal{F}$。

根据德摩根定律可以推导出性质(ⅴ)。有的时候也把性质(ⅴ)作为域的定义条件之一，但实际上这是没有必要的，因为性质(ⅰ)和(ⅱ)就足够了。

例 2-19 最简单的(但是平凡的)域为 $\mathcal{F}=\{\phi,\Omega\}$。 ◀

例 2-20 单次硬币投掷试验的样本空间为 $\Omega=\{H,T\}$，考虑非平凡的域 $\mathcal{F}=\{\phi,\Omega,H,T\}$。上面有关域的特性可以在此验证：(ⅰ)$H\in\mathcal{F}\Rightarrow T\in\mathcal{F}$，反之亦然。(ⅱ)～(ⅲ)$H\in\mathcal{F}$，$T\in$

$\mathcal{F}\Rightarrow H\cup T\equiv\Omega\in\mathcal{F}$。(iv)～(v)$H\in\mathcal{F}$，$T\in\mathcal{F}\Rightarrow H\cap T\equiv\phi\in\mathcal{F}$。实际上，$\mathcal{F}$是这个试验的幂集$\mathcal{P}(\Omega)$。除了例 2-19 中的那个平凡域，其他任何域都不可能是幂集。◀

在有些试验中，我们首先考虑一个并不包含域的事件集合，之后再补充一些事件产生一个域。我们通过例 2-21 来描述这个过程。

例 2-21 假设样本空间 $\Omega=[1,\infty)$。对于任意的 $a>1$，$\mathcal{F}=\{\phi,\Omega,\{[1,a)\}\}$，那么$\mathcal{F}$包含所有$[1,a)$形式的半开半闭区间。很显然$\mathcal{F}$并不是一个域，因为对于任意的 $a>1$，区间 $[1,a)^c=[a,\infty)$ 并不在$\mathcal{F}$中。将集合扩展为 $\mathcal{F}=\{\phi,\Omega,\{[1,a)\},\{[b,\infty)\}\}$，$b>1$。同样的，$\mathcal{F}$也不是一个域，因为对于任意的 $b<a$，区间 $[1,a)\cap[b,\infty)=[b,a]$ 并不在$\mathcal{F}$中。再次将集合扩展为 $\mathcal{F}=\{\phi,\Omega,\{[1,a)\},\{[b,\infty)\},\{(c,d)\}\}$，所有的 $a,b,c,d>1$。此时得的一个域，因为所有事件的有限并集都是$\mathcal{F}$中的事件，也是这些事件的补集。在接下来的例题中，我们将利用无限并集来定义一个$\mathcal{F}$，此时的$\mathcal{F}$并不是 σ 一个域。◀

例 2-22 对于两次硬币投掷试验，样本空间 $\Omega=\{HH,TT,HT,TH\}$。表 2-3 归纳了幂集$\mathcal{P}(\Omega)$中的所有元素。在这里，性质(i)和(ii)很容易得到验证。从这个例题可以看出很有必要总结一下事件的性质。表 2-3 的第一行给出了试验的全部四种输出结果，它们都是基本事件，该表格其他行给出的都不是基本事件。比如，$\{HH,HT\}$ 表示第一枚硬币朝向为 H 的事件；$\{HH,TT\}$ 表示两枚硬币朝向一致的事件。这两个事件都不是输出结果本身，它们给出的是具有共同特征的输出结果。$\{HH,HT,TH\}$ 表示输出结果中至少有一个 H 的事件，三个输出的组合就更难描述了。事件 $\{HH,TT,HT\}$ 就不太好叙述，这个事件表示两枚硬币的朝向一致，或者第一枚硬币是 H 第二枚硬币是 T 的情况。对于这个事件还可以有另外一种方法描述，第一枚硬币不是 T 的输出，以及第二枚硬币不是 H 的输出(因此将 TH 的情况排除掉了)。因为我们要求$\mathcal{F}$中所有事件的并集都应该在域中，因此 Ω 也必须在$\mathcal{F}$中。同样的，$\phi\in\mathcal{F}$表示对于$\mathcal{F}$中的任意一个事件，其补集和交集都是一致的。可以看出$\mathcal{F}$中有 16 个元素，可知 $N=4$，幂集的基数为 $|\mathcal{P}(\Omega)|=2^4$。◀

表 2-3　例 2-22 域$\mathcal{F}=\mathcal{P}(\Omega)$的所有元素

基本事件(四种输出)	$\{HH\},\{TT\},\{HT\},\{TH\}$
根据两种输出定义的事件	$\{HH,TT\},\{HH,HT\},\{HH,TH\},\{TT,HT\},\{TT,TH\},\{HT,TH\}$
根据三种输出定义的事件	$\{HH,TT,HT\},\{HH,TT,TH\},\{HH,HT,TH\},\{TT,HT,TH\}$
ϕ 与 Ω	$\phi=\Omega^c,\Omega=\{HH,TT,HT,TH\}$

根据定义幂集$\mathcal{P}(\Omega)$是最大的那个域。从前面可知，域并不是唯一的，这点从例 2-23 也可以看出。一般说来，对于有限或无限可数输出的试验而言，较小的域并没有太多用处。但是对于无限不可数的试验，我们就非常有必要挑选出比幂集小的 σ 域。

例 2-23 从例 2-19 可知，$\{\phi,\Omega\}$ 是一个域，它完全满足性质(i)和(ii)。很显然，对于任何试验而言，这个平凡域几乎没有任何用处。现考虑掷单个骰子的试验，输出结果为 $\{1,2,3,4,5,6\}$。幂集$\mathcal{P}(\Omega)$包含了 $2^6=64$ 个元素。但是，有可能定义一个更小的(非平凡的)域，尽管它的用处没有幂集大。考虑两个事件：$E=\{1,3,5\}$ 和 $F=\{2,4,6\}$，以及域 $\mathcal{F}=\{\phi,\Omega,E,F\}$。很显然，性质(i)和(ii)是满足的，以及 $E=F^c\in\mathcal{F}$，$E\cup F=\Omega\in\mathcal{F}$。尽管这也是一个有效的域，但它实际上是这个试验的一个简化版本，此时我们只关心输出结果是奇数还是偶数，并不关心骰子上具体的点数是多少。实际上，通过定义这么一个域，这个“简化的”试验其实与单次投掷硬币试验是完全等价的，只不过这个硬币的 H 面表示的是“偶数”，T 面表示的是“奇数”。根据上面这个较小的域的定义，投掷骰子就可以模拟投掷硬币试验。◀

请再次思考这个情况：首先定义 Ω 的子集，然后再产生一个域包含这个子集。

例 2-24 对于单次投掷骰子试验，首先根据子集$\{1,2\}$产生一个域。$\{1,2\}^c=\{3,4,5,6\}$，$\{1,2\}\cap\{3,4,5,6\}=\phi$，$\{1,2\}\cup\{3,4,5,6\}=\Omega$，因此根据这个子集产生的域为$\mathcal{F}=\{\phi,\Omega,\{1,2\},\{3,4,5,6\}\}$。很显然，这个域并不能包含原始试验的所有输出。实际上，这等同于一个二元试验，比如单次硬币投掷试验(见例 2-23)。再考虑子集$\{\{1,2\},\{3,4\}\}$，根据这个子集产生的域为$\mathcal{F}=\{\phi,\Omega,\{1,2\},\{3,4\},\{5,6\},\{3,4,5,6\},\{1,2,5,6\},\{1,2,3,4\}\}$。为了能够把单次投掷骰子试验的所有结果都包含进去，我们就需要选择包含$2^6=64$个元素的幂集来生成域。◀

上面最后一个例题给我们描述了一个“好”域与一个“差”域的区别，前者由幂集定义，后者由幂集的子集定义。在后续学习事件的条件概率的时候，我们将用“好”域与“差”域的区别来推导出概率论中一个非常有用的性质。

接下来，我们将根据原子的定义来描述域$\mathcal{F}$的一个性质，全集Ω是有限的。

定义(域的原子) 域F中最小的事件(除了ϕ)称作**原子**。我们用$\mathcal{A}$来表示原子的集合。

根据这个定义可知，原子不能通过$\mathcal{F}$中其他事件的并集获得，而且它们是不相交的。$\mathcal{F}$中其他所有的事件都是通过对$\mathcal{A}$中的元素进行集合运算得到的(这样是为了满足域的条件)。图 2-4 给出了一个简单的例子，它形象地表示了Ω的输出，$\mathcal{A}$中的原子以及$\mathcal{F}$中事件的相互关系。

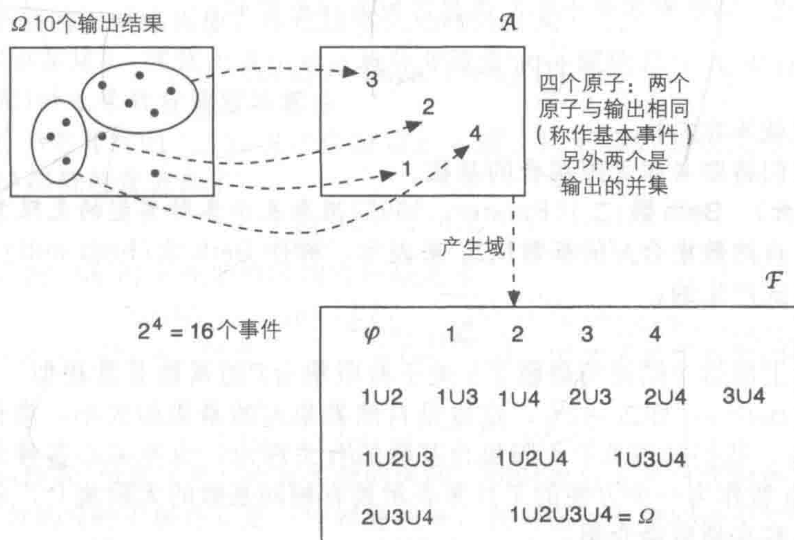


图 2-4 Ω中的输出，$\mathcal{A}$中的原子以及$\mathcal{F}$中事件的相互关系

例 2-25 对于例 2-24 的第二个域，三个原子分别是$\mathcal{A}=\{\{1,2\},\{3,4\},\{5,6\}\}$，但它们并不是$\Omega$的输出。◀

很显然，一个试验幂集的原子就是它的基本事件(输出)。如果把域中独立的输出都排除掉，那么例 2-24 中的原子都不是基本事件。

定理 2-1 任意有限域$\mathcal{F}$中元素的数目为2^n，其中$n=|\mathcal{A}|$。

证明：根据原子的定义以及域的要求可知：$\mathcal{F}$的元素一定是原子的幂集$\mathcal{P}(\mathcal{A})$。对于一个含N个元素的集合，其幂集的基数为2^N，因此有$N=n$，证毕。

定理 2-1 非常有用，因为我们可以据此马上得出结论：如果$\mathcal{F}$中元素的数目不是 2，4，8，16，那么$\mathcal{F}$一定不是一个域，以此类推。请注意，并不需要Ω是离散的来保证域是有限的，在此利用例 2-26 来进行讲解。

例 2-26 设$\Omega=[0,\infty)=\mathcal{R}^+$，$\mathcal{F}=\{\phi,\Omega,[0,c),[c,\infty)\}$，$c>0$。尽管这个样本空间是连续的，域只包含 4 个元素，其原子由两个子区间$[0,c)$和$[c,\infty)$给出。如果假设原子为$[0,c)$，$[c,d)$和$[d,\infty)$，并且$d>c>0$。根据上面的定理可知，这个连续样本空间的域包

含 8 个元素，即 $\mathcal{F}=\{\phi,\Omega,[0,c),[c,d),[d,\infty),[0,d),[c,\infty),[0,c)\cup[d,\infty)\}$。实数域 $\mathcal{R}$上的这种区间对于连续样本空间非常重要，我们在后续介绍波雷尔(Borel)集合的时候还将进一步讨论。◀

对于有限样本空间的问题，根据条件(ⅰ)和(ⅱ)给出的域的定义等同于定义一个概率空间。但是，如果这个样本空间是无限的，不管它是无限可数还是不可数，此时就有必要将域的定义进行扩展，使其包含一个无限(可数)事件的并集。

定义(σ-域)　**σ-域**既要满足条件(ⅰ)，还要满足扩展的条件(ⅱ)，即将条件(ⅱ)扩展到无限可数个子集的情况，即：

(ⅱ′) 如果 $A_n\in\mathcal{F}$，$n=1,2,\cdots$，那么 $\bigcup_{n=1}^{\infty}A_n\in\mathcal{F}$。

σ-域也称作 σ-代数。

很显然，上面这个修改的定义也将条件(v)进行了扩展，具体为：

(ⅴ′) 如果 $A_n\in\mathcal{F}$，$n=1,2,\cdots$，那么 $\bigcap_{n=1}^{\infty}A_n\in\mathcal{F}$。

例 2-27 根据例 2-21，当 $a,b,c,d>1$，$\mathcal{F}=\{\phi,\Omega,\{[1,a)\},\{[b,\infty)\},\{(c,d)\}\}$ 为 $\Omega=[1,\infty)$ 的一个域，但它不是一个 σ-域。假设将 $[c,d]$ 写为 $[c+1/n,d]$ 的形式，n 为正整数。根据表 2-2，以及前面关于不同类型区间的讨论可知：

$$\bigcup_{n=1}^{\infty}[c+1/n,d)=(c,d) \tag{2-18}$$

这种类型的区间就不在$\mathcal{F}$中。◀

接下来，我们将要考虑无限集合的基数。

定义(Beth 数)　**Beth 数**$\{\beth_n\}$(Forster，1995)用来表示各种类型的无限集合的基数。

在数学中，自然数集合$\mathcal{N}$的基数用$\beth_0$ 来表示，称作 Beth 零(beth null)。Beth 数是按照如下的递归方式产生的，

$$\beth_{n+1}=2^{\beth_n} \tag{2-19}$$

可以看出，上面这个结论与命题 2-1 关于有限集合$\mathcal{P}$的基数非常相似。这样就可知实数集$\mathcal{R}$的基数为 Beth 1，即$\beth_1=2^{\beth_0}$，这也是自然数集$\mathcal{N}$的幂集的大小。类似的，$\mathcal{P}(\mathcal{R})$的基数为 Beth 2 $\beth_2$。我们并不关心无限集合基数的有关理论，也不关心各种无穷大的特性。我们只是将 Beth 数作为一个方便的工具来表示具有相同基数的无限集合。表 2-4给出了一些例题，有些内容在稍后会介绍。

表 2-4　Beth 数与无限集合的基数

beth 数	具有相同基数的无限集合
$\beth_0$	$\mathcal{N}$，$\mathcal{Q}$，$\mathcal{Z}$
$\beth_1$	$\mathcal{R}$，$\mathcal{R}-\mathcal{Q}$，$\mathcal{P}(\mathcal{N})$，$\mathcal{B}(\mathcal{R})$，$\mathcal{C}$，$[a,b]\subset\mathcal{R}$，康托尔集
$\beth_2$	$\mathcal{P}(\mathcal{R})$，$\mathcal{P}(\mathcal{P}(\mathcal{N}))$

对于$\Omega=\mathcal{N}$，其幂集$\mathcal{P}(\mathcal{N})$为一个 σ-域。实际上，任意 Ω 的幂集都是σ-域，而且从前面的讨论易知：任意 Ω 的σ-域都满足$\{\Omega,\phi\}\subseteq\mathcal{F}\subseteq\mathcal{P}(\Omega)$。

定义(事件空间(Event Space))　事件空间为 Ω 中组成σ-域$\mathcal{F}$的所有事件的集合。事件空间写作$\{\Omega,\mathcal{F}\}$。

事件空间包含了满足 σ-域条件的所有事件。我们将例 2-21 和例 2-24 中如何产生一个域的思想扩展到 σ-域。

定义(产生一个 σ-域)　对于子集$\mathcal{G}\subseteq\Omega$，定义 σ-域的集合如下：

$$E=\{\mathcal{F}\subseteq\mathcal{P}(\Omega):\mathcal{G}\subseteq\mathcal{F}\} \tag{2-20}$$

通过子集$\mathcal{G}$产生的σ-域可由E中所有事件取交集得到，记作$\sigma(\mathcal{G})$。

因为Ω中的每个σ-域都包含$\mathcal{G}$，因此$\sigma(\mathcal{G})$必然是包含$\mathcal{G}$的最小σ-域。如果$\mathcal{G}$本身就是一个σ-域，那么$\sigma(\mathcal{G})=\mathcal{G}$。

例 2-28 试考虑有限样本空间$\Omega=\{-1,0,1\}$。下面给出所有可能的σ-域：

$$\mathcal{F}_1=\{\phi,\Omega,-1,0,1,\{-1,0\},\{-1,1\},\{0,1\}\}, \tag{2-21}$$

$$\mathcal{F}_2=\{\phi,\Omega,-1,\{0,1\}\}, \tag{2-22}$$

$$\mathcal{F}_3=\{\phi,\Omega,0,\{-1,1\}\} \tag{2-23}$$

$$\mathcal{F}_4=\{\phi,\Omega,1,\{-1,0\}\}, \tag{2-24}$$

$$\mathcal{F}_5=\{\phi,\Omega\} \tag{2-25}$$

如果$\mathcal{G}=\Omega$，那么$\sigma(\mathcal{G})=\mathcal{F}_1\cap\mathcal{F}_2\cap\mathcal{F}_3\cap\mathcal{F}_4\cap\mathcal{F}_5=\mathcal{F}_5$为一个平凡$\sigma$-域。如果$\mathcal{G}=\{1\}$，那么$\sigma(\mathcal{G})=\mathcal{F}_1\cap\mathcal{F}_4=\mathcal{F}_4$。如果$\mathcal{G}=\{0,1\}$，那么$\sigma(\mathcal{G})=\mathcal{F}_1$为幂集$\mathcal{P}(\Omega)$。◀

我们感兴趣的是那些可测的σ-域，这样就可以按照类似的方法给出事件空间$\{\Omega,\mathcal{F}\}$的概率。对于有限和无限可数的试验，没有必要考虑比幂集$\mathcal{P}(\Omega)$小的σ-域。但是对于连续的样本空间，幂集显得“太大”了。本章最后介绍的 Vitali 集合(此时$\Omega=[0,1]$)就是一个子集不可测的例子。对于绝大多数的实际问题和应用而言，用区间就可以很方便的表示连续试验。从第 3 章可知，样本空间都映射成为了实线集合$\mathcal{R}$上的随机变量，因此我们对$\mathcal{R}$上一种特定类型的σ-域感兴趣，称作波雷尔(Borel)σ-域。

定义(波雷尔σ-域) 实线上最小的σ-域称作**波雷尔 σ-域**$\mathcal{B}(\mathcal{R})$，它由$\Omega=\mathcal{R}$上所有的开区间生成。$\mathcal{B}(\mathcal{R})$的元素称作**波雷尔集合**。

利用$\mathcal{R}$上的一个开区间(a,b)来产生波雷尔σ-域。从下面的表达式可知，所有的孤立点(单元素集合)都被包含进去了：

$$a=\lim_{n\to\infty}(a-1/n,a+1/n) \tag{2-26}$$

这样也可以把所有的半开半闭区间也包括进来：

$$(a,b]=(a,b)\cup b,\quad [a,b)=(a,b)\cup a \tag{2-27}$$

因为$(-\infty,b]=(b,\infty)^c$以及$[a,\infty)=(-\infty,a)^c$，因此可知所有的闭区间也包含进来了：

$$[a,b]=[a,\infty)\cap(-\infty,b] \tag{2-28}$$

因此可知，$\mathcal{B}(\mathcal{R})$包含了$\mathcal{R}$上所有类型的区间，同时也包含了全部的独立点。想找出那些位于$\mathcal{B}(\mathcal{R})$之外的$\mathcal{R}$的子集并不是一件容易的事，在后面讨论测度理论的时候，我们给出这种子集的一个例子。由于实际应用场合并不会出现这种子集，因此我们对其并不关心。在第 4 章介绍随机向量的时候，我们还将考虑$\Omega=\mathcal{R}^N$上的σ-域，因此需要将前面的定义扩展到$\mathcal{R}^N$子集中最小的σ-域，记作$\mathcal{B}(\mathcal{R}^N)$，并将其定义为$\mathcal{R}^2$上的开矩形，或者是$\mathcal{R}^N$上的高维开矩形($N>2$)。

例 2-29 波雷尔集合的例子包括：(i)闭区间$[a,b]$，(ii)开区间(a,b)，(iii)有理数集$\mathcal{Q}$，(iv)无理数集$\mathcal{R}-\mathcal{Q}$，(v)$[1,2,3,4,5,6]$，(vi)后面将要介绍的康托尔集。◀

2.5 随机试验的总结

在此有必要将随机试验的有关性质进行总结归纳，这样我们就可以进行概率测度：

- 一个随机试验所有输出结果的集合为样本空间Ω。根据试验的不同类型，它可能包含有限、无限可数或无限不可数的元素。
- 一个随机试验的事件是样本空间Ω的子集，它包含一个或多个输出，并具有某些共同的特性。我们的目的是计算事件的概率测度，因此我们并不愿意包含Ω所有的子集。
- σ-域$\mathcal{F}$是Ω子集的集合，并满足代数条件(i)，(ii′)，(iii)，(iv)和(v′)。可以通过Ω的一些子集来构造一个σ-域，然后将子集的数量进行扩展直到满足条件(i)和(ii′)。

- 如果Ω是有限或无限可数的，那么通过幂集$\mathcal{P}(\Omega)$描述的所有子集中一定包含一个对于该试验有用的σ-域。
- 如果Ω是无限不可数的，那么波雷尔σ-域$\mathcal{B}(\mathcal{R})$是实线上一个有用的σ-域，它包含了所有类型的区间(包括开区间，半开半闭区间，闭区间)以及独立点。

根据样本空间Ω和事件空间$\{\Omega, \mathcal{F}\}$的定义，我们可以得到一组事件的集合，该集合满足(ⅱ′)中的叠加性质。至此我们可以给出概率空间的定义，记作$\{\Omega, \mathcal{F}, \mathcal{P}\}$。首先我们简要介绍一下测度理论，其中概率测度是一个特别的例子。

2.6 测度理论

测度理论关心的是如何给样本空间Ω中的子集赋予一个"尺寸"，我们称之为测度。根据问题的不同，这个尺寸有可能是长度、面积或者欧式空间的体积，通过后面的定义可知这些统称L(Lebesgue，勒贝格)测度。在本书中，我们对如何给Ω中那些组成σ-域$\mathcal{F}$的子集赋予概率测度感兴趣。对于不可数的试验，用类似的方法不可能给$\mathcal{P}(\mathcal{R})$赋予概率。$\Omega=\mathcal{R}$的幂集太大了(它的基数为 Beth 2 $\beth_2$)，所以此时我们选择用$\mathcal{F}$来表示$\mathcal{B}(\mathcal{R})$，它是实线上的波雷尔σ-域，基数为 Beth 1 $\beth_1$(与$\mathcal{R}$相同)。在许多实际应用场合波雷尔σ-域是非常有用的，因为我们比较喜欢用实线上的区间来描述事件。

用符号μ来表示事件空间中的测度，用$\{\Omega, \mathcal{F}, \mu\}$来表示测度空间。(请注意不要将这里的$\mu$与第5章的随机变量均值相混淆。)

定义(测度) **测度**是指事件$\{E_n\}\in\mathcal{F}$到$\mathcal{R}$的一种映射，并满足如下性质：

- 对于所有的事件$E_n\in\mathcal{F}$，都有$\mu(E_n)\geqslant 0$。
- $\mu(\phi)=0$。
- 对于所有的互斥事件，即当$n\neq m$时$E_n\cap E_m=\phi$，

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty}\mu(E_n) \tag{2-29}$$

在后面介绍概率测度的时候也有类似的性质，称之为概率三原则(the three axioms of probability)。

例 2-30 如果事件空间$\{\Omega, \mathcal{F}\}$中的事件$E_n\in\mathcal{F}$都满足$\mu(E_n)=0$，则为平凡测度。◀

例 2-31 对于$\Omega=\mathcal{R}$和$\mathcal{F}=\mathcal{B}(\mathcal{R})$，标准高斯测度为：

$$\mu(E) = \frac{1}{\sqrt{2\pi}}\int_E \exp(-x^2/2)\mathrm{d}x \tag{2-30}$$

这即为事件E的概率值$P(E)$，E是零均值，方差为1的标准高斯随机变量。◀

例 2-32 对于事件空间$\{\Omega, \mathcal{F}\}$中的事件E，其狄拉克测度为：

$$\delta_\omega(E) \triangleq \begin{cases} 1, & \omega\in E \\ 0, & \text{其他} \end{cases} \tag{2-31}$$

容易验证$\delta_\omega(E)$是一个概率测度，因此E是Ω中的几乎确定性事件(见习题2-23)。◀

下面两个重要的测度，在后面将会用到。

定义(计数测度(Counting Measure)) 设Ω为一个可数的样本空间，$\mathcal{P}(\Omega)$是幂集。任意一个集合$E\in\mathcal{P}(\Omega)$，其计数测度为：

$$C(E) = \begin{cases} |E|, & E\text{是有限子集} \\ \infty, & \text{其他} \end{cases} \tag{2-32}$$

$|E|$表示集合E的基数。在上面这个定义中，无穷大其实就是前面提到的 Beth 零$\beth_0$，也是无限不可数集合的大小。

很显然，测度只对有限集合才有意义，对于无限集合(比如$\mathcal{R}$上的区间)测度不能提供任何信息。对于后一种情况，我们将考虑L测度。

定义（$\mathcal{R}$上的勒贝格测度）　对于区间$[a,b]\in\mathcal{B}(\mathcal{R})$，其 L **测度**定义如下：

$$L([a,b])\triangleq b-a \tag{2-33}$$

该定义表示区间的**长度**。根据定义，单个点（单元素集合）的 L 测度为 0，因此可得 $L((a,b))=L((a,b])=L([a,b))=b-a$。

为了将 L 测度的定义扩展 N 维欧式空间，在此给出笛卡儿乘积的定义。

定义（笛卡儿乘积(Cartesian product)）　一组区间集合 $\{[a_n,b_n]\},n=1,\cdots,N$ 的**笛卡儿乘积**定义如下：

$$E\triangleq\{(x_1,\cdots,x_N):x_1\in[a_1、b_1],\cdots,x_N\in[a_N,b_N]\} \tag{2-34}$$

E 对应于一个有序 N 元组，可以用下面的乘积符号来表示：

$$E=[a_1,b_1]\times\cdots\times[a_N,b_N]=\prod_{n=1}^{N}[a_n,b_n] \tag{2-35}$$

$\mathcal{R}$区间的笛卡儿乘积，对应于$\mathcal{R}^2$ 的矩形，在$\mathcal{R}^N(N>2)$上是超矩形。

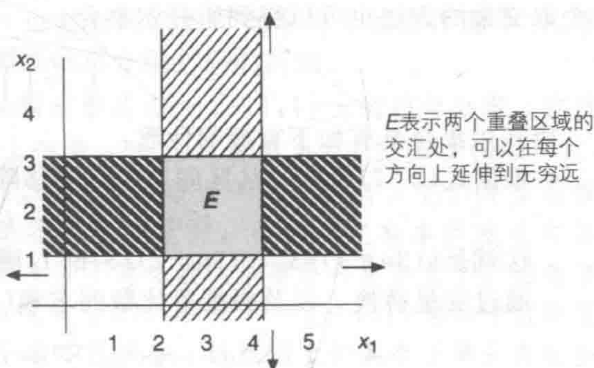


图 2-5　例 2-33 中 $E=[2,\ 4]\times[1,\ 3]$的笛卡儿乘积

例 2-33　对于区间$[1,3]$和$[2,4]$，笛卡儿乘积 $E=\{(x_1,x_2):x_1\in[2,4],x_2\in[1,3]\}$是$\mathcal{R}^2$ 上的一个矩形，该区域由横坐标在[2, 4]之间所有点与纵坐标在[1,3]之间所有点的交集构成，如图 2-5所示。◀

对于实线集合$\mathcal{R}$，我们可以写成$\mathcal{R}^2=\mathcal{R}\times\mathcal{R}$(平面)，$\mathcal{R}^3=\mathcal{R}\times\mathcal{R}\times\mathcal{R}$(超平面)的形式，以此类推。从几何学的观点看，$E$ 的 L 测度对应着一个矩形区域的面积，以及一个三维空间的体积。对于 N 维空间，我们需要给出如下的定义。

定义（$\mathcal{R}^N$ 上的勒贝格测度）　笛卡儿乘积 $E=\prod_{n=1}^{N}[a_n,b_n]\in\mathcal{B}(R^N)$ 的 L **测度**定义为：

$$L(E)\triangleq\prod_{n=1}^{N}(b_n-a_n) \tag{2-36}$$

该定义表示超矩形(hyper-rectangle) $\mathcal{R}^N$ 的超体积(hyper-volume)。

例 2-34　例 2-33 中的笛卡儿乘积 $L([1,3]\times[2,4])=4$ 表示的是一个矩形区域的面积，而 $L([8,10]\times[1,2]\times[4,8])=8$ 表示的是一个立方体的体积。◀

在例 2-35 中介绍的康托尔集合非常有趣，一是因为该集合的构建方式，另外是因为它的 L 测度为 0。

例 2-35　(康托尔集合)在区间$[0,1]$上去除一个如下所示的无限开区间，可得到康托尔集合 C(Kingman and Taylor，1966)。

$$C=[0,1]-\bigcup_{m=1}^{\infty}\bigcup_{n=0}^{2^{m-1}-1}((3n+1)/3^m,(3n+2)/3^m) \tag{2-37}$$ ◀

设 $C_0=[0,\ 1]$，C_m 代表连续的集合，$m\in\mathcal{N}$。将 C_0 等分为三部分，去除中间的那段开区间，可得 C_1 为：

$$C_1=[0,1]-(1/3,2/3)=[0,1/3]\cup[2/3,1] \tag{2-38}$$

此时可以看出区间$[0,1]$被等分成了三部分，中间那段开区间被移除了。对 C_1 中剩下的每段闭区间都进行如此操作，可得：

$$C_2=C_1-(1/9,2/9)-(7/9,8/9)=[1,1/9]\cup[2/9,1/3]\cup[2/3,7/9]\cup[8/9,1] \tag{2-39}$$

图 2-6 给出了每个 C_m 的求解过程。从式(2-37)可以看出，每个连续集合都是通过上一个集合求出的，即首先将区间除以 3，然后再将这个缩放后的区间平移 2/3，最后对这两部分取并集。下面的这个公式刻画了整个处理流程：

$$C_m = [C_{m-1}/3] \cup [C_{m-1}/3 + 2/3] \tag{2-40}$$

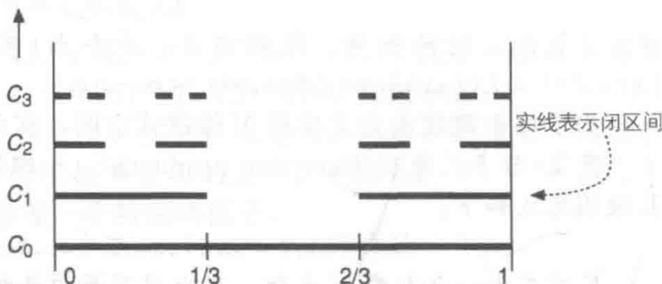


图 2-6 $m=1, 2, 3$ 时 C_m 的序列，当 $m\to\infty$ 时得到康托尔集合 C

这个过程就是不停地将区间中间 1/3 的部分移除，一直进行下去就可以得到 C。需要注意的是在式(2-37)中移除的都是开区间，而利用下面的这个取交集的方法也可以得到康托尔集合：

$$C = \bigcap_{m=0}^{\infty} C_m \tag{2-41}$$

康托尔集合具有如下有趣的性质：

- 根据式(2-37)可知，从区间 [0,1] 中移除的每个开区间的长度为：
$$(3n+2)/3^m - (3n+1)/3^m = 1/3^m \tag{2-42}$$
这就是$((3n+1)/3^m, (3n+2)/3)$的 L 测度。对于每个 m，有 2^{m-1}个开区间被移除。通过变量替换，以及根据等比数列求和(见附录 E)，整个被移除的区间长度为：
$$\sum_{m=1}^{\infty} 2^{m-1}/3^m = \sum_{m=0}^{\infty} 2^m/3^{m-1} = 3/(1-2/3) = 1 \tag{2-43}$$
这个结论非常令人惊讶，因为区间 [0,1] 的长度(L 长度)也是 1。这是因为在构建$\{C_m\}$的每次迭代过程中，中间区间的端点并没有移除，而实际上，无论一个区间的端点是否被包括在内，它的 L 测度都是相同的。
- 根据前面的结论可知，C 的 L 测度为 0。
- C 是不可数的，它的基数与同区间 [0,1] 相同(与$\mathcal{R}$的基数相同，记作 Beth1 $\beth_1$)。
- C 是分形的，因为在每次迭代后，构建出的 C_{m+1}都是前一个 C_m 的压缩版。在图 2-6 中，前面几次迭代中也有相同的特性。

在后面的例 3-10 中，将介绍如何根据康托尔集合定义康托尔函数。

L 测度在第 3 章中非常有用，比如将 Ω 中的事件映射到实线$\mathcal{R}$上得到随机变量，或者映射到$\mathcal{R}^N$ 上得到随机向量。因为 Ω 中的事件到随机变量的映射具有明确的定义(可测的)，这些事件的概率测度就会以一种连续的方式对应于$\mathcal{R}$区间上的事件(以及$\mathcal{R}^N$ 上的超矩形)。

最后，我们给出样本空间 Ω 上事件概率测度的定义。

定义(概率测度 P) 事件空间$\{\Omega, \mathcal{F}\}$的**概率测度** P 是一个函数，该函数将$\mathcal{F}$中的事件映射为区间 [0,1] 中的实数，该定义满足概率论三大公理。得到的概率空间记作$\{\Omega, \mathcal{F}, \mathcal{P}\}$。

需要特意强调的是，P 是为 σ-域$\mathcal{F}$中的事件定义的，而$\mathcal{F}$不一定包含 Ω 中所有的子集。当然这就是构建$\mathcal{F}$的目的，此时概率测度就可以与事件对应上，并且满足概率三大公理。与任意一种测度类似，概率测度必须满足可数的叠加条件才显得有意义，当对$\mathcal{F}$中事件进行集合运算时，才能与我们关于概率计算的直观感受相符合。

2.7 概率公理

概率空间$\{\Omega, \mathcal{F}, P\}$的概率测度 P 满足下面的公理：

公理 2-1 $P(E)\geqslant 0$。

公理 2-2 $P(\Omega)=1$。

公理 2-3 当 $E_n \cap E_m = \phi$ 且 $n \neq m$ 时，对于 $\{E_m\} \in \mathcal{F}$：

$$P\left(\bigcup_{m=1}^{\infty} E_m\right) = \sum_{m=1}^{\infty} p(E_m) \tag{2-44}$$

上面的这三大概率公理与我们的直观感受是相符的，也与第 1 章中关于概率的频率解释一致。公理 2-1 表明概率测度是非负的。公理 2-2 表明 $\mathcal{F}$ 中某些事件的发生概率是 1："必然发生的事件"。根据公理 2-1 和 2-2，$\mathcal{F}$ 中任意事件的概率一定位于区间 $[0,1]$ 之间，任何事件的概率不可能超过 1。P 是一般测度 μ 的一个特例，$\mu(\phi)=0$ 改用 $P(\Omega)=1$ 表示。这个区别就是公理 2-2 中 $P \in [0,1]$ 的原因，所有的测度 μ 必须是非负的，但并不要求不大于 1，而概率测度就有这个限制。

$P(E)=0$ 并不表示事件 E 不会发生，一个事件的概率为 0 与不可能事件是有区别的，前者是可以发生的，而后在是绝对不会发生的。同样的，$P(E)=1$ 并不意味着事件 E 一定会发生，一个事件的概率为 1 与必然事件也是有区别的。概率为 1 的事件称作"几乎确定性事件(almost sure event)"。下面这个例题就将介绍它们的区别。

例 2-36 对于一个随机事件，它的可能输出服从 $\Omega = [-1,1]$ 上的均匀分布。在后面的讨论可知，对于连续样本空间，该区间上任意一个点的概率为 0，比如 $P(E)=0$。但是，事件 $E=0$ 是有可能发生的。区间 $[-1,1]$ 之外的就是不可能事件，它们发生的概率为 0，并且是绝对不会发生的。空集 ϕ 也是不可能发生的，因为一定有事件发生在区间 $[-1,1]$ 之内。类似的，$P(F=[-1,1))=1$ 并不意味着这个事件一定会发生，因为事件 $E=1$ 也是有可能发生的(尽管事件 E 发生的概率为 0)。事件 $F=[-1,1)$ 是几乎肯定发生的。另外一方面，事件 $F=[-1,1]$ 是一个确定性事件，该区间上的某个子集是肯定会发生的，因为不可能出现该区间之外的数字。◀

公理 2-3 是这三大公理中最重要的，它确保了概率测度的一致性，它也指出了如何将多个事件的概率整合在一起。如果这些事件是对偶互斥的(不相交)，那么它们并集的概率就等于它们各自概率之和。这也适用于无限多事件的并集，对于 σ-域中的无限可数和不可数试验是非常需要的。在下面的章节，我们将介绍概率论中的一些基本结论，包括条件概率(conditional probability)和独立事件(independent event)这两个重要知识点。后面我们还将介绍如何计算离散和连续样本空间中的概率，这部分内容为第 3 章中将要介绍的随机变量做准备。

2.8 概率论的一些基本结论

下面的这些命题来自于概率论三大公理。通过它们我们可以了解如何以集合不相交的观点来重新组织问题，这些结论通过公理 2-3 也是可以推导出的。

命题 2-2 $P(\phi)=0$。

证明：因为 $\Omega \cup \phi = \Omega$ 以及 $\Omega \cap \phi = \phi$，可得：

$$P(\Omega) = P(\Omega \cup \phi) = P(\Omega) + P(\phi) \tag{2-45}$$

可知 $1 = 1 + P(\phi)$，$P(\phi)=0$。

命题 2-3 对于事件 $E \in \Omega$，$P(E^c) = 1 - P(E)$。

证明：因为 $\Omega = E \cup E^c$ 以及 $E \cap E^c = \phi$，可得：

$$P(\Omega) = P(E^c + E) = P(E^c) + P(E) = 1 \tag{2-46}$$

上式可以改写为 $P(E^c) = 1 - P(E)$。

命题 2-4 如果 $E \subset F$，那么 $P(E) \leqslant P(F)$。

证明：因为 F 包含 E，F 就可以用 E 来进行划分：

$$F = E \cup E^c F \tag{2-47}$$

由于所有的概率值都是非负的，因此可得：

$$P(F) = P(E) + P(E^c F) \geqslant P(E) \tag{2-48}$$

命题 2-5 是本节最重要的内容，它可以看做是公理 2-3 关于不相交事件的一个拓展。

命题 2-5　对于事件 E，$F\in\Omega$，有下面结论成立：

$$P(E\cup F)=P(E)+P(F)-P(EF) \tag{2-49}$$

证明：尽管事件 E 和 F 可能是不相交的，但可以将 $E\cup F$ 写成三个互斥事件并集的形式：

$$E\cup F=EF^c\cup E^cF\cup EF \tag{2-50}$$

这个关系如图 2-7 所示。根据概率论的公理 2-3可知：

$$P(E\cup F)=P(EF^c)+P(E^cF)+P(EF) \tag{2-51}$$

将事件 E 和 F 写成不相交事件并集的形式：$E=EF^c\cup EF$，$F=E^cF\cup EF$，由此可得：

$EF^c\cup EF\cup E^cF=E\cup F$

$EF^c\cap EF=\Phi$

$EF^c\cap E^cF=\Phi$

$EF\cap E^cF=\Phi$

E　F　Ω　EF^c　EF　E^cF

图 2-7　将 $E\cup F$ 拆分成三个不相交的事件

$$P(E)=P(EF^c)+P(EF),\quad P(F)=P(E^cF)+P(EF) \tag{2-52}$$

将上面的结论代入式(2-51)可得：

$$P(E\cup F)=P(E)-P(EF)+P(F)-P(EF)+P(EF)=P(E)+P(F)-P(EF) \tag{2-53}$$

命题得证。

例 2-37　现考虑单次硬币投掷试验。设事件 $E=\{偶数\}$，$F=\{点数<3\}$。很显然这二者并非不相交事件。因为 $P(E)=1/2$，$P(F)=1/3$，以及 $P(EF)=1/6$，根据式(2-49)可得：

$$P(E\cup F)=1/2+2/3-1/6=2/3 \tag{2-54}$$

图 2-8 用维恩图的形式很清晰的解释了这个结论。◀

E表示输出值是偶数

F表示输出值小于3

$E\cap F$=表示输出值是偶数

并且小于3，即$E\cap F=\{2\}$

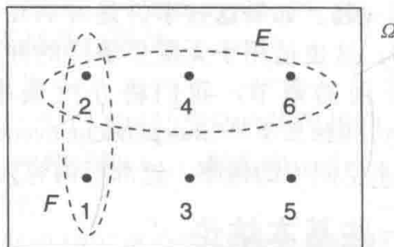


图 2-8　例 2-37 中，利用维恩图来计算事件的概率，这些事件不一定是不相交的

2.9　条件概率

在许多场合，随机事件中有些事件的信息是已知的，而且该部分事件可以影响到其他事件发生的概率。可以将这部分先验信息用于计算事件的概率。对于概率空间 $\{\Omega,\mathcal{F},P\}$ 的事件 A 和 B，如果已知事件 B 已经发生，那么事件 A 发生的概率往往与事件 B 未知的情况不同(尽管这不是绝对的)。常用符号 $P(A|B)$来表示 B 已知时 A 的概率(也称作已知事件 B 为真)。接下来，我们用一个简单例题来介绍如何计算条件概率。

例 2-38　对于投掷两枚硬币的试验，设事件 A 表示两枚硬币都是正面朝上，事件 B 表示至少有一枚硬币正面朝上。这个例子就可以很轻松的解释条件事件(比如，可以获得额外的信息)是如何将样本空间从四个元素压缩到三个元素。因为事件 TT 被排除了，事件 B 使得“新”的样本空间是 $\{HH,HT,TH\}$，因此我们可以马上计算出 $P(A|B)=1/3$。我们同样也可以计算 $P(B|A)$，尽管这个概率没有什么意义。此时压缩后的样本空间仅仅为$\{HH\}$，很显然事件 B(至少有一枚硬币正面朝上)始终为真，因此可得 $P(B|A)=1$。◀

$P(A|B)$就是一个条件概率。在例 2-38 中，根据压缩后的样本空间使得条件概率可以

比较容易地计算出来，但在一般情况下我们使用基于事件交集的定义。

定义(条件概率) 事件 B 发生时事件 A 的**条件概率**为：

$$P(A|B) \triangleq \frac{P(AB)}{P(B)} \tag{2-55}$$

在此假设 $P(B)\neq 0$。

图 2-9 用维恩图的形式很好地解释了事件 A 和 B 关于条件概率的定义。当 B 为真时，原始样本空间 Ω 被压缩为只包含 B 的元素的那部分(因为此时事件 B 已经发生了)。因此，式(2-55)中的分子表示 A 和 B 中共有的元素：因为 B 为真，我们只对 A 和 B 交集中那部分元素的概率感兴趣。计算 $P(AB)$ 需要依据的是原始样本空间中的所有原始输出结果，但样本空间此时已经被压缩了，因此需要将 $P(AB)$ 进行归一化处理。将 $P(AB)$ 除以 $P(B)$，使得 $P(\cdot|B)$ 是一个恒定大小的概率测度，并且满足概率论的公理，下面的这个命题将对其进行证明。

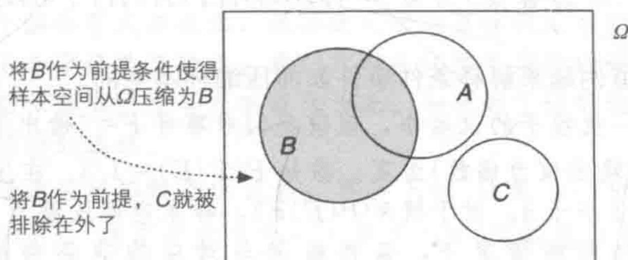


图 2-9 条件概率：如果元素属于 A 而不属于 B，那么在测度 $P(\cdot|B)$ 中它们的概率为 0 (实际上，它们就是不可能事件)

命题 2-6 条件概率 $P(A|B)$ 是一个关于事件 A 的有效概率测度。

证明：根据公理 2-1 和条件概率的定义可知，$P(AB)/P(B)\geqslant 0$。又因为 $P(AB)$ 和 $P(B)$ 是原始样本空间 Ω 中有效的概率，立即可知 $P(A|B)\geqslant 0$。根据公理 2-2，

$$P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \tag{2-56}$$

最后，根据公理 2-3 以及无限可数不相交事件$\{A_n\}$可知，

$$P\left(\bigcup_{n=1}^{\infty} A_n|B\right) = \frac{1}{P(B)}P\left(\bigcup_{n=1}^{\infty} A_nB\right) = \frac{1}{P(B)}\sum_{n=1}^{\infty}P(A_nB) = \sum_{n=1}^{\infty}P(A_n|B) \tag{2-57}$$

在证明最后，关于对联合事件$\{A_nB\}$取并集用到了公理 2-3 的结论。

对于有限试验，$P(B)=|B|/|\Omega|$，此时我们还可以根据概率的频率解释来证明这个命题是正确的。根据压缩后的样本空间：

$$P(A|B) = \frac{|AB|}{|B|} \tag{2-58}$$

上面这个公式可以很好的描述例 2-38 中的条件概率是如何计算的。同时请注意这个等式还可以改写为：

$$P(A|B) = \frac{|AB|/|\Omega|}{|B|/|\Omega|} = \frac{P(AB)}{P(B)} \tag{2-59}$$

得到的结果同定义是相同的。尽管我们不能将概率的频率解释推广到连续试验，但直觉告诉我们这应该是相同的，而且定义(2-55)可以应用到任何“好定义”的随机试验：比如连续的、离散的或者混合的。

需要注意的是，$P(A|B)$只是关于事件 A 的有效概率测度，关于条件事件 B 并不是有效的。对于事件 A，B 和 C，下面的表达式是命题 2-6 的一个合理拓展：

$$P(A\cup B|C) = P(A|C) + P(B|C) - P(AB|C) \tag{2-60}$$

条件事件 C 被移到了每一项的右端。如果将条件事件写成集合运算的形式，比如

$P(A|B\cup C)$，那么一般而言，

$$P(A|B\cup C)\neq P(A|B)+P(A|C)-P(A|BC) \tag{2-61}$$

对于条件事件进行集合运算不满足概率论公理，而且我们不能将关于独立事件并集的条件概率拆开成各自条件概率的形式，比如不能将 $P(A|B\cup C)$写成 B 和 C 的分开的表达式。

再次考虑原始的样本空间 $\{\Omega,\mathcal{F},P\}$。如果将 B 作为前提条件，那么样本空间被压缩成了 $B\subset\Omega$，σ-域也被压缩为了子 σ-域$\mathcal{G}\subset\mathcal{F}$，条件概率为 $P(\cdot|B)$，与前面描述的一样。因此，此时的条件概率空间为 $\{B,\mathcal{G},P(\cdot|B)\}$，接下来我们将用一道例题来进行讲解。

例 2-39 投掷两枚硬币试验的样本空间为 $\Omega=\{HH,TT,TH,HT\}$。设$\mathcal{F}$为幂集，其元素见表 2-3。定义条件事件 $B=\{$至少有一个是 $H\}$，样本空间被压缩为 $B=\{HH,TH,HT\}$，对应的子 σ-域（B 的幂集）为 $\mathcal{G}=\{\phi,B,HH,TH,HT,\{HH,TH\},\{HH,HT\},\{TH,HT\}\}$。 ◀

我们再用另外一道例题来解释条件事件如何压缩样本空间。

例 2-40 在投掷一枚骰子的试验中，假设我们对事件 $F=\{$输出值$<3\}$的概率感兴趣，并且前提是事件 $E=\{$输出值为偶数$\}$为真。易知 $P(F|E)=1/3$。在压缩后的样本空间 $\{2,4,6\}$ 中，只有输出值 2 小于 3。对于概率 $P(E|F)$，样本空间压缩为 $\{1,2\}$，我们马上可知 $P(E|F)=1/2$。在这两种情况下，条件概率与对应的非条件概率值都是一样的：$P(F|E)=P(F)$以及 $P(E|F)=P(E)$。这不是一个巧合，这是因为在这个试验中事件 E 和 F 是相互独立的。 ◀

2.10 独立性

独立性是概率论中一个非常重要的特性，在随机试验的建模中，该特性往往可以使分析更加简便。如果两个事件对各自的输出结果没有任何影响，那么我们就可以认为这两个事件是相互独立的。具体定义如下：

定义（独立事件） 事件 A 和 B 是相互**独立**的，当且仅当下面的等式成立：

$$P(AB)=P(A)P(B) \tag{2-62}$$

$P(AB)$称作事件 A 和 B 的联合概率（joint probability，表示二者同时发生）。在联合概率的概念里，独立概率 $P(A)$和 $P(B)$称作边缘概率（marginal probability）。接下来的这个例题就很清晰的讲解了独立性的定义。

例 2-41 投掷两枚硬币，观察到的试验结果都是正面朝上的概率为 $P(HH)=P(H)P(H)=(1/2)(1/2)=1/4$。很显然，从物理学的角度来看，一枚硬币的投掷不可能影响到另外一枚硬币投掷的结果，从概率学的频率角度解释，这个概率就可以拆分成式(2-62)那样的独立事件的形式。同样的，投掷两枚骰子，观察到任意一对输出结果 $\{\omega_m,\omega_n\}$ 的概率为 $P(\omega_m,\omega_n)=P(\omega_m)P(\omega_n)=1/36$，其中 $\omega_m\in\{1,2,3,4,5,6\}$。 ◀

根据条件概率的定义可知：

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B)}{P(B)}=P(A) \tag{2-63}$$

同样的，

$$P(B|A)=P(B) \tag{2-64}$$

从上面这两个结论看出，两个事件的独立性是一个对称的特性，即二者是相互独立的。式(2-62)给出的是独立性的标准定义，式(2-63)也可以当做独立性的定义。实际上，后者的结论更直观一些，因为事件 B 作为事件 A 的前提也无法影响 A 的概率，很显然它们是独立的。为了验证两个事件是否独立，我们只需要验证式(2-55)的左边与右边是否相等即

可。也就是说，如果等式两边的数值是相等的，那么事件是独立的。非独立事件无法保证式(2-62)成立。

两个事件独立性的定义可以拓展到多个事件。

定义(相互独立事件(Mutually Independent Events)) 事件$\{A_1,\cdots,A_N\}$**相互独立**，是指下式对于$\{A_1,\cdots,A_N\}$所有可能的子集成立：

$$P(A_m\cdots A_n)=P(A_m)\cdots P(A_n) \tag{2-65}$$

这不仅包括两个事件的组合，也包括三个事件的组合，以此类推，适用于N个事件组合的情况：

$$P(A_mA_n)=P(A_m)P(A_n),\quad m\neq n \tag{2-66}$$

$$P(A_mA_nA_p)=P(A_m)P(A_n)P(A_p),\quad m\neq n\neq p \tag{2-67}$$

$$\vdots$$

$$P(A_1\cdots A_N)=P(A_1)\cdots P(A_N) \tag{2-68}$$

如果上面任意一个组合等式不成立，根据定义这些事件就是非独立的。通过这种方法可以比较容易的判断多个事件是否独立。我们可以先从两个事件的组合开始验证，然后验证三个的情况，一直验证到有一个等式不成立为止。如果为了证明这些事件都是相互独立的，所有的等式都必须得到验证。

例 2-42 还是继续前面投掷骰子的试验，事件$E=\{$偶数$\}$，$F=\{$小于 3$\}$，我们知道此时$P(EF)=P(E)P(F)$。现在考虑另外一个事件$G=\{$可以被 3 整除$\}$，此时对应的是$G=\{3,6\}$，而且$P(G)=1/3$。根据图 2-8 的维恩图可知，$EG=\{6\}$，因此$P(EG)=1/6$。可以看出$P(FG)=0$(因为F和G是不相交的)，但是$P(F)P(G)=1/9$。事件$\{E,F,G\}$不是相互独立的。◀

2.11 贝叶斯公式

贝叶斯公式在计算条件概率，或者从条件概率中计算边缘概率时非常有用。

定理 2-2(贝叶斯公式) 对于事件$\{A,B\}\in\Omega$，且$P(B)\neq 0$：

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \tag{2-69}$$

同样地，对于$P(A)\neq 0$：

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)} \tag{2-70}$$

证明：根据条件概率的定义，我们可以写成下面的表达式：

$$P(AB)=P(B|A)P(A)=P(A|B)P(B) \tag{2-71}$$

整理上面这个结论，$P(A|B)$或$P(B|A)$的表达式为贝叶斯公式，也称作贝叶斯法则。

例 2-43 (二元对称信道)。贝叶斯公式最重要的一个例子就是二元对称信道(BSC)，如图 2-10 所示，它给出了一个数字通信系统中发生比特传送错误的模型。对于一个二元系统，被传输的符号有两种输出，我们用$\{0,1\}$来表示这两个二元基本事件。这个模型其实与硬币投掷试验是等效的，只需要定义$H\equiv 1$和$T\equiv 0$。在第 3 章将要介绍，用随机变量X来表示被传输的符号是很方便的，此时X"随机"的取这两个值。X本身不是随机的：它被两个可能的输出值(称作字母表)定义，而且每个取值都赋予一个特定的概率。随机变量的随机性体现在：之前谁也无法知道到底会取哪个值，X的

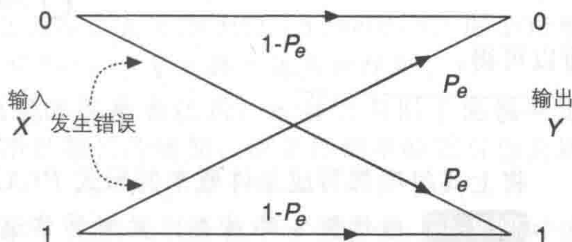


图 2-10 二元对称信道(BSC)，输入为X，输出为Y。"二元"意味着$|X|=|Y|=2$，"对称"意味着$P(Y=0|X=0)=P(Y=1|X=0)\triangleq P_e$

取值是不确定的。一般情况下，在一个真实系统中 $P(X=0)=P(X=1)=1/2$(类似于硬币投掷试验)。

同样的，用随机变量 Y 来表示接收到的符号，它的取值同样是字母表 $\{0,1\}$。我们感兴趣的是在这个信道里发生符号传输错误的概率，也可以用两个条件概率来表示这种错误，即 $P(Y=1|X=0)$和 $P(Y=0|X=1)$。如果这两种条件概率是相等的，那么这个信道是对称的。可以用关于 X 的条件概率来定义平均错误概率：

$$P_e \triangleq P(Y=1|X=0)P(X=0)+P(Y=0|X=1)P(X=1) \tag{2-72}$$

这就是关于全概率定律的一个例子，稍后我们会进行讨论。因为信道是对称的，因此只需要计算一个条件概率，可得：

$$P_e = P(Y=1|X=0)[P(X=0)+P(X=1)] = P(Y=1|X=0) \tag{2-73}$$

在一个实际的通信系统中，可以通过很多次试验来测量 P_e 的数值，一般会发送一长串的符号 0 和 1，而且接收端事先知道发送的符号，最后统计错误比特数。

通信系统的接收端需要相反的条件概率：$P(X=0|Y=0)$，$P(X=0|Y=1)$，$P(X=1|Y=0)$，以及 $P(X=1|Y=1)$。例如，如果收到的符号是 $Y=0$，接收端的任务就是判断到底发送的是 $X=0$ 还是 $X=1$。因为输入是随机的，接收端无法完全准确的判断传送的是哪个符号。有一种判决准则称作最大后验概率准则(MAP，见第 10 章)，该准则给出的结果会使得 $P(X=x|Y=y)$取最大值，其中 $x,y \in \{0,1\}$。根据贝叶斯公式，还可以写为如下形式，

$$P(X=x|Y=y)=\frac{P(Y=y|X=x)P(X=x)}{P(Y=y)} \tag{2-74}$$

其中 $P(Y=y|X=x)$对于一个特定的通信系统是已知的。接收端事先知道先验概率 $P(X=x)$(在以事件 Y 为前提之前)，因此 $P(Y=y)$可以通过下面的全概率公式计算得到：

$$P(Y=y)=P(Y=y|X=0)P(X=0)+P(Y=y|X=1)P(X=1) \tag{2-75}$$

在第 2.12 节中还将继续这个例题。 ◀

2.12 全概率

全概率定律作为贝叶斯公式的一个拓展，将样本空间 Ω 划分为了不同的事件。在有些问题中利用全概率定律来计算某个事件的概率会比较方便，尤其是涉及条件概率的情况，例如例 2-43 讨论的二元对称信道就属于这种情况。

定理 2-3(全概率) 设 $\{B_n\}, n=1,\cdots,N$，是样本空间 Ω 的一个划分，那么下面的等式成立：

$$P(A)=\sum_{n=1}^{N}P(A|B_n)P(B_n) \tag{2-76}$$

证明：因为$\{B_n\}$来自一个划分，所以 $B_n \cap B_m=\phi$ 以及 $AB_n \cap AB_m=\phi$，$n\neq m$。又因为：

$$\bigcup_{n=1}^{N} AB_n=\Omega \tag{2-77}$$

所以可得：

$$P(A)=\sum_{n=1}^{N}P(AB_n) \tag{2-78}$$

将上式每项都写成条件概率的形式 $P(AB_n)=P(A|B_n)P(B_n)$，命题得证。

例 2-44 继续例 2-43 中的二元对称信道，输入为 X，输出为 Y，对 X 进行划分可以得到式(2-75)，每一个 Y 的取值概率为：

$$P(Y=0)=P(Y=0|X=0)P(X=0)+P(Y=0|X=1)P(X=1) \tag{2-79}$$

$$P(Y=1)=P(Y=1|X=0)P(X=0)+P(Y=1|X=1)P(X=1) \tag{2-80}$$

先验概率$\{P(X=x)\}$和信道错误概率$\{P(Y=y|X=x)\}$往往是已知的，可以计算得到$\{P(Y=y)\}$。设$\{P(X=1)\}=p$，因为信道是对称的，可以将 $P(Y=1|X=0)=P(Y=$

$0|X=1)=P_e$ 和 $P(Y=1|X=1)=P(Y=0|X=0)=1-P_e$ 代入式(2-79)和(2-80)得到：

$$P(Y=0)=(1-P_e)(1-p)+P_e p \tag{2-81}$$

$$P(Y=1)=P_e(1-p)+(1-P_e)p \tag{2-82}$$

图 2-11 给出了上面两个关于 p 的函数的图像，此时 $P_e=0.1$（对于一个通信系统而言这是一个很高的错误概率）。很显然，当 $p\neq 1/2$ 时，$P(Y=0)\neq P(Y=1)$（$p=1/2$ 是指 $P(X=0)=P(X=1)$）。

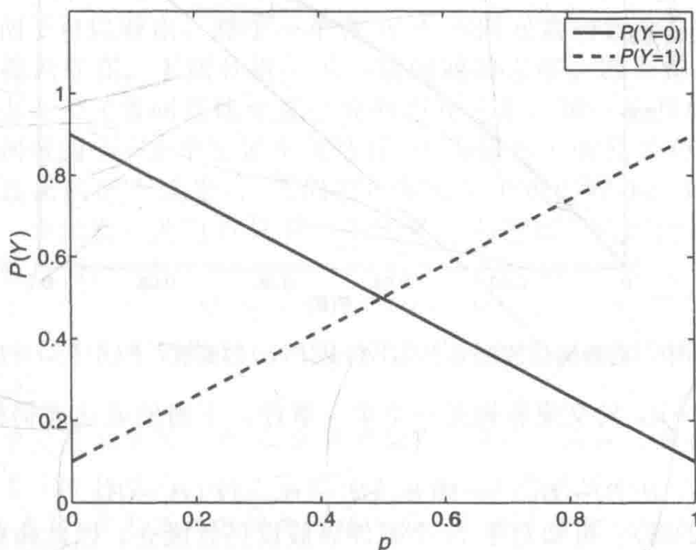


图 2-11　例 2-44 中，二元对称信道的输出概率，该概率是 $p=P(X=1)$ 的函数，此处的错误概率 $P_e=0.1$ ◀

例 2-45（药物测试）。药物测试是有关贝叶斯公式和全概率定律的另外一个非常有名的应用，这个测试用于判断病人是否患病。两个事件：$A=\{$测试呈阳性$\}$，$B=\{$病人患病$\}$。存在两种类型的错误：$P(A|B^c)$ 即假阳性，$P(A^c|B)$ 即假阴性。在研制药物的过程中，这两个概率可以通过试验结果估算得到。此外，基于已有的大量数据，我们假设 $P(B)$ 是已知的，并且准确程度很高。因此，当一个病人患病并且得出的测试结果为阳性的概率就可以用贝叶斯公式计算为：

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)} \tag{2-83}$$

根据全概率定律可知：

$$P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)} \tag{2-84}$$

等式右边的所有项都是已知的，因为 $P(A|B)+P(A^c|B)=1$，$P(A|B^c)+P(A^c|B^c)=1$。例如，如果 $P(A^c|B)=P(A|B^c)=0.1$，$P(B)=0.02$，那么可知 $P(B|A)\approx 0.16$。图 2-12 给出了 $P(B|A)$ 的函数图像，横坐标的范围是 0 到 0.1，表示人群中患病的概率。◀

最后，联合概率可以写成一个有关条件概率的连乘表达式，这有点类似于全概率定律，不过这里并不需要对事件进行划分。随着事件数目的增加，取条件概率的数目也会增加，下面的这个定理将进行介绍。

定理 2-4　设样本空间 Ω 中的 N 个事件 $\{B_n\}$，它们的联合概率 $P(B_1\cdots B_N)\neq 0$。对这些事件连续取条件概率可得，

$$P(B_1\cdots B_N)=P(B_1)P(B_2|B_1)P(B_3|B_1B_2)\cdots P(B_N|B_1\ldots B_{N-1}) \tag{2-85}$$

证明：通过数学归纳法来证明。根据贝叶斯公式，对一个事件取条件概率为：

$$P(B_1B_2)=P(B_1)P(B_2|B_1) \tag{2-86}$$

假设对于 $N-1$ 个事件，下面的表达式成立：

$$P(B_1\cdots B_{N-1}) = P(B_1)P(B_2\,|\,B_1)P(B_3\,|\,B_1B_2)\cdots P(B_{N-1}\,|\,B_1\cdots B_{N-2}) \tag{2-87}$$

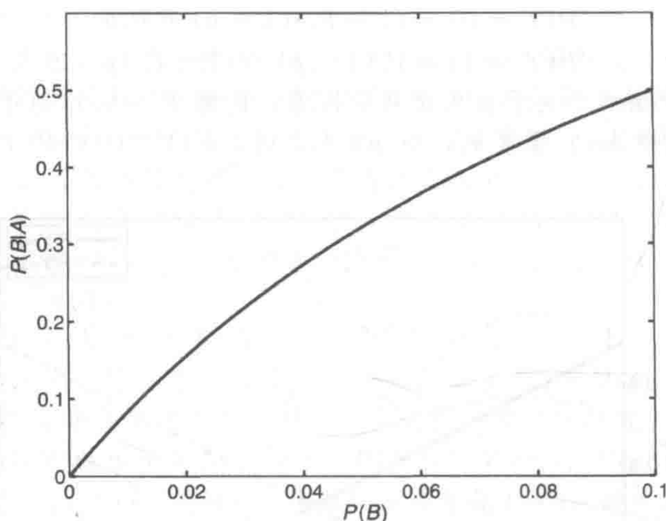


图 2-12　例 2-45 中，药物测试的概率 $P(B\,|\,A)$是 $P(B)$的函数，$P(A^c\,|\,B)=P(A\,|\,B^c)=0.1$

因为可以把 $B_1\cdots B_N$ 的交集看做是一个单一事件，下面的表达式仍然是贝叶斯公式的一种形式：

$$P(B_1\cdots B_N) = P(B_N\,|\,B_1\cdots B_{N-1})P(B_1\cdots B_{N-1}) \tag{2-88}$$

将式(2-87)代入式(2-88)，可知对于 N 个事件该假设仍然成立，因此命题得证。

2.13　离散样本空间

给离散样本空间(有限或无限可数)赋予概率值是一种给很直观的做法，因为在样本空间 Ω 中事件被当作了独立的样本点，并且在赋予概率值时满足概率论公理。同前面类似，在此我们假设离散样本空间 Ω 的幂集为$\mathcal{F}=\mathcal{P}(\Omega)$。

例 2-46 某随机试验的输出$\{x_n\}$为自然数集$\mathcal{N}$。假设我们对这些输出赋予如下的概率值：

$$P(x_n) = (1-p)^{n-1}p \tag{2-89}$$

其中 $0<p<1$。这种类型的概率得到的是几何分布随机变量(见第 3 章)。实际上，这对应的就是硬币投掷试验，将试验连续进行直到第一次出现 H 为止，其中 $P(H)=p$。尽管输出结果有无数多个，但全部概率之和为 1：

$$\sum_{n=1}^{\infty}(1-p)^{n-1}p = p\sum_{n=0}^{\infty}(1-p)^n = \frac{p}{1-(1-p)} = 1 \tag{2-90}$$

上面的结论用到了附录 E 中关于无限长序列求和的闭合表达式。这是一个收敛序列的例子，该序列被 p 进行了比例缩减，同时这是一个可数长度的序列，因此该序列求和结果收敛为 1。实际上，任意收敛序列只要其求和结果为 1，那么该序列都可以被赋予概率。◀

接下来考虑一个特殊的有限样本空间 Ω，其输出结果$\{\omega_n\}$是等概率的。(很显然，无限可数的样本空间不可能存在等概率的输出)。在这种情况下，任意一个事件 $E\in\Omega$ 的概率等于事件 E 发生的次数与样本空间基数的比值：

$$P(E) = \frac{|E|}{|\Omega|} \tag{2-91}$$

其中$|E|$表示 E 的基数。在这个等概率的情况下，概率的计算简化为了一个组合问题。计算$|\Omega|$相对比较简单，计算$|E|$要复杂一些，因为这涉及输出的子集。为了对这个问题有更清晰的认识，我们需要学习组合与排列之间的区别。

定义(组合与排列(Combination and Permutation))　**组合**是指不同元素的无序集合。

排列是指这些不同元素的一种有序的集合。

对于 N 个不同的元素，只存在一种组合：因为不用考虑元素的顺序，随便什么顺序都是相同的组合。而另外一方面，对这些元素不同的顺序就对应了不同的排列。

例 2-47 对于集合 $E=\{a_1,a_2,a_3\}$，只有一种可能的组合，即使 $\{a_2,a_1,a_3\}$ 也是同一个组合。另外一方面，存在6种可能的排列：除了前面提到的两种，还有 $\{a_1,a_3,a_2\}$，$\{a_2,a_3,a_1\}$，$\{a_3,a_1,a_2\}$ 和 $\{a_3,a_2,a_1\}$。◀

从这个简单的例子可以看出，对于一个含 N 个不同元素的集合，只可能有一种组合存在，但有 $N!$ 种排列存在。下面分析一下一般问题的求解。因为排列是一个有序的集合，意味着需要考虑一个元素同其他元素位置的相互关系，第一步我们有 N 种可能的选择，第二步只能在剩余的 $N-1$ 个元素里面选择(因为排列中的元素必须是不同的，我们不能选择第一步中选出的那个元素)。类似的，在第三步我们有 $N-2$ 个元素可以选择，以此类推直到最后一个元素，此时就只剩一个元素可供选择了。因此，对于 N 个不同的元素，可能的排列数目为：

$$N(N-1)\cdots 2\cdot 1=N! \tag{2-92}$$

如果要统计事件 E 中输出的数目，我们必须首先考虑它们的顺序是否重要。接下来的例题描述了这种情况。

例 2-48 假设我们要从桌面上的52张牌中抽出5张不同的牌。那么存在5！种可能来整理这些牌，对应着120种排列。但是我们知道，在扑克牌游戏中，手里面牌的排列并不能影响你的“牌面实力”。我们抓到的牌的好坏与否只取决于牌的组合，它们的顺序是无关紧要的。(即使我们拿到的是同花顺 $\{5,6,7,8,9\}$：我们将它们按照升序排列还是降序排列，也不会影响这手牌的实力。)◀

例 2-49 一般情况下，如果在问题描述中着重强调了元素的摆放，那么就应该考虑这是一个排列问题。现考虑社保号(social security number，SSN 是一个9位数字，并且这9位数字是不同的，不包括0)的数目统计问题。很显然，这些数字的顺序就很重要了：存在着 $9!=362880$ 个可能的社会安全号，也即是数字1到9的排列问题。这种类型的“试验”称作“非替代性采样”(sampling without replacement)：如果一个数字被选中了，那么该数字就不会被再次选中。如果假设我们允许数字被重复选用，此时就会有 9^9 种可能(仍然不包括0)。后面这种情况就不属于式(2-92)的范畴，因为这些数字不是相互不同的，这称作“替代性采样”(sampling with replacement)。此时元素个数的计算与推导式(2-92)的过程比较类似：允许数字被重复选用的话，就有会有 $9\times 9\times\cdots\times 9=9^9=387420489$ 种可能。◀

对于涉及 N 个不同但等概率输出结果的问题，我们往往对其某个子集感兴趣，该子集对应某个事件且包含 $M<N$ 个元素。这种情况下存在不止一种组合，因为即使不考虑顺序，从 N 个元素中选出 M 个元素来也有很多种方法。易知，这种情况下排列的数目为：

$$N(N-1)\cdots(N-M+1)\triangleq(N)_M \tag{2-93}$$

其中 $(N)_M$ 表示递降阶乘(falling factorial)(见附录B)，它可以被写作如下形式：

$$(N)_M=\frac{N(N-1)\cdots(N-M+1)}{(N-M)\cdots 2\cdot 1}[(N-M)\cdots 2\cdot 1]=\frac{N!}{(N-M)!} \tag{2-94}$$

因为这个集合有 M 个元素，因此根据式(2-92)可知存在着 $M!$ 种排列。计算从 N 个元素中选出 M 个元素的组合数目，将式(2-94)除以排列的数目即可。也就是把含有相同 M 个元素的所有排列当做是同一种组合。

$$\frac{N!}{(N-M)!M!}=\frac{(N)_M}{M!}\triangleq\binom{N}{M} \tag{2-95}$$

上面最后一个表达式称作二项系数，我们读作“从 N 个中选择 M 个”。式(2-94)和式(2-95)分别总结了从 N 个不同元素中选取 M 个元素的组合数目与排列数目。在附录E的表E-1和E-2中给出了一些阶乘和二项式系数结果，并给出了Stirling公式，用于估算 N 很大时的 $N!$。

例 2-50 假设我们对6位车牌号码感兴趣(不考虑字母)，并且这些数字不能重复，也不能包括数字0。根据式(2-94)，$N=9$，$M=6$，此时数字的顺序是很重要的，因此车牌的数目为一个排列问题：

$$\frac{9!}{(9-6)!}=504 \tag{2-96}$$

如果我们不考虑数字的顺序，根据式(2-95)，车牌的数目为一个组合问题：

$$\binom{9}{6}=84 \tag{2-97}$$

如果按照组合问题看待，车牌123456与213456就是“相同的”，因此它们就只能算一个车牌。根据式(2-95)，将结果除以6个数字的排列数目($M!=6!=720$)即可。◀

上面的结论可以归纳成这样一个试验：假设存在 N 个元素但这些元素并非相互不同，将这些元素归拢到 M 个子集中去，每个子集中的元素都是相同的。第一组有 N_1 个完全相同的元素，第二组有 N_2 个完全相同的元素，以此类推，第 M 组有 N_M 个完全相同的元素，并且有：

$$\sum_{m=1}^{M} N_m = N \tag{2-98}$$

前面的例题也符合这个试验，只不过每一组里面只有一个元素：$N_M=1, m=1,\cdots,M$。(请注意：很显然，我们假设组与组之间不含有相同的元素，否则的话，这些组就应该被归拢到一个更大的组里面去。)

因为这 N 个元素不再是相互区别的，原有的排列数目 $N!$ 就随着这些组的排列数目而降低。第一个组里面有 N_1 个完全相同的元素，很显然需要将 $N!$ 除以 $N_1!$，这样第一组里面所有的排列都算作一个。以此类推，可得总的排列数为：

$$\frac{N!}{N_1!N_2!\cdots N_M!} \triangleq \binom{N}{N_1,N_2,\cdots,N_M} \tag{2-99}$$

等式右边的表达式称作多项式系数。如果所有的组都只有一个元素，式(2-99)就简化为了式(2-92)。

例 2-51 考虑一个投掷4枚硬币的随机试验，设事件 E 表示出现3个反面。可知 E 的元素为 $\{T,T,T,H\}$，$\{T,T,H,T\}$，$\{T,H,T,T\}$ 和 $\{H,T,T,T\}$，这也是3个反面1个正面的所有可能的排列。因此 $|E|=4$，$|\Omega|=2^4=16$，概率 $P(E)=|E|/|\Omega|=1/4$。在这个试验中元素的顺序很重要：我们必须考虑出现这一个 H 的4种方式。在这里不能直接应用式(2-92)，因为这里的元素并不是相互区别的：出现的3个反面是相同的元素，这3个 T 在事件 E 的4个元素中顺序并不重要。根据式(2-99)，$N=4$，$N_1=1$，$N_3=3$，可得：

$$|E|=\binom{4}{3,1}=4 \tag{2-100}$$

接下来再考虑事件 F。在 F 中包含两个 T，我们再分析一下问题又会有什么变化。此时 F 的元素为 $\{T,T,H,H\}$，$\{H,H,T,T\}$，$\{H,T,H,T\}$，$\{H,T,T,H\}$ 和 $\{T,H,H,T\}$。根据式(2-99)，F 的组合数目为，

$$|F|=\binom{4}{2,2}=6 \tag{2-101}$$

因此概率 $P(F)=|F|/|\Omega|=3/8$。◀

最后，我们进一步讨论对有限样本空间的采样，且结果为等概率输出。再次考虑前面的两个例题：（ⅰ）在扑克游戏中手里抓到的5张牌；（ⅱ）投掷一枚骰子。对于第一个例子，这5张牌是从桌面上抽出来的，它们并没有相互替代，这是一个非替代性采样的例子。这道例题一个显著的特点就是：如果还要从同一个桌面上抽出5张牌（比如另外一个玩家），那么样本空间已经发生了变化，此时只有47张牌可供选择。在计算概率的时候这种情况肯定要考虑到。第二个例子对应的是替代性采样：每次投掷骰子的时候，样本空间都是没有变化的。在每次投骰子后6个点数都是可能的。如果将这个骰子投掷 M 次，排列的数目为：

$$\underbrace{N\cdots N}_{M} = N^M \tag{2-102}$$

对于一个有限的随机试验，表2-5总结了各种情况下样本数目的计算公式。其中第四种情况对应的是习题2-49，它给出了如何计算可替代时的排列数目。

表2-5 排列与组合的数目

采样类型（N 个元素选出 M 个）	公式
非替代性组合	$\binom{N}{M}=\frac{N!}{M!\ (N-M)!}$
非替代性排列	$(N)_M=\frac{N!}{(N-M)!}$
替代性组合	$\binom{N+M-1}{M}=\frac{(N+M-1)!}{M!\ (N-1)!}$
替代性排列	N^M

例2-52 假设将一个骰子投掷两次。根据表2-5给出的公式可知，可替代时的组合数目为：

$$\binom{6+2-1}{2}=\frac{7!}{2!5!}=21 \tag{2-103}$$

考虑所有可能的输出结果，根据第一个骰子的结果将它们分成如下的组：

$$\begin{aligned}&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\\&(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\\&(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\end{aligned} \tag{2-104}$$

上面给出的就是所有的可替代性排列，一共有 $N^M=6^2=36$ 个。上面的每个结果都是唯一的，因为在计算排列的时候顺序是很重要的。对于组合顺序就不重要了，第二组里面的 (2,1) 和第一组里面的 (1,2) 就是同一种组合。同样的，我们还需要将 $\{(3,1),(3,2)\}$，$\{(4,1),(4,2),(4,3)\}$，$\{(5,1),(5,2),(5,3),(5,4)\}$ 和 $\{(6,1),(6,2),(6,3),(6,4),(6,5)\}$ 分别从其他四个集合里面排除掉。将这15个输出结果从原始的36种排列里面扣除掉，得到的就是21个替代性的组合，同时也验证了式(2-103)的结果。◀

尽管不一定真的要以掷骰子的方式选择一个输出，但我们仍可以将这个试验看作是对样本空间 Ω 的一种采样。这两种类型的采样最大的区别在于：在计算概率的时候，替换性采样的样本空间 $|\Omega|$ 不会发生变化。假设有一个连续时刻的采样，如果这个采样是可替换的，那么采样输出是相互独立的，因为前一个采样并不会影响接下来的采样。如果是非替换的采样，此时情况就不一样了：后续的样本概率还需要考虑之前有哪些样本从样本空间 Ω 中被移走了。

本节讨论了有限离散样本空间。对于无限可数的离散空间，采用第3章中将要介绍的随机变量的概率质量函数(pmf)会更加方便。在本章最后，我们将介绍连续样本空间，在第3章中还将详细介绍这部分内容。

2.14 连续样本空间

连续样本空间的输出结果是不可数的，这种情况下不可能给每个独立样本点赋予一个非零的概率值，并且这些概率值之和为 1。例如，一个随机试验的输出为非负实数，$\Omega=\mathcal{R}^+$。很显然，任意两个非负实数之间总有无限多个非负实数(该结论不适用于输出结果为可数的随机试验)。如果我们试图给独立点赋予概率值，那我们很快认识到这些概率值不可能分配给$\mathcal{R}^+$中所有的数并同时满足概率论公理。如果一个随机试验的输出是连续的，那么取到一个具体数值的概率必定为 0。这个结论乍一看上去似乎有点违背常识，因为在一个连续试验中，取到某个具体数值肯定是必然的，但通过某些示例可以看出这个结论的正确性。比如，取到 $\pi=3.14159265\cdots$这个是“几乎不可能的”，这个事件基本上是不会发生的，换种说法即 $P(\pi)=0$。

给连续随机试验赋予概率值，对于实线空间$\mathcal{R}$上需要考虑的是区间，对于$\mathcal{R}^2$ 而言是矩形，$\mathcal{R}^N$ 上就是*超矩形*。基于区间的思路来赋概率值，不仅可以满足概率论公理，而且也符合很多工程实际问题。例如一个反映日均气温的随机问题，我们往往会对气温在 70—75 ℉(华氏度)之间的概率感兴趣，而不会对一个特定的气温值(比如 72.354 ℉)感兴趣。将问题聚焦为事件的一个区间，这样就将一个不可数的试验“转换”为了一个可数试验，并且可以用可数样本空间的基本方法对其进行处理，同时还满足概率论的可数叠加公理(公理 2-3)。

例 2-53 考虑如图 2-13 所示的随机试验，样本空间 $\Omega=[0,1]$，我们对三个子区间 $\{E,F,G\}$ 的概率感兴趣。可以看出除了这三个子区间，还有其他 4 个子区间共同组成了样本空间，即 $\Omega=[0,1]$ 被划分成了 7 个子区间，给这些区间赋概率值满足概率论公理。同样的，每个子区间还可以被进一步划分，只要不被划分到一个个的孤立点，这些区间都是可以被赋概率值的。最后一点需要指出的是：如果我们想对孤立点赋非零的概率值，那么这应该是一个混合随机试验，输出结果既包含离散的，也包含连续的成分。在第 3 章中讨论映射到不同类型随机变量的时候将介绍这部分内容。 ◀

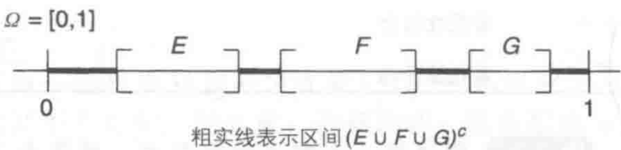


图 2-13 连续样本空间中的一段区间被赋非零的概率值，孤立点的概率为 0。通过区间，样本空间 Ω 被划分为了可数的(有限多)事件

对于实线上任意区间的一个集合，我们总能在它们中间找到可数多个区间，使得所有区间的并集组成样本空间 $\Omega=\mathcal{R}$。这就可以解释为什么波雷尔 σ-域在概率空间 $\{\mathcal{R},\mathcal{B}(\mathcal{R}),P\}$ 中会非常重要，它包括实线上所有的开区间、闭区间、半开半闭区间，以及孤立点。在许多实际应用中，我们不仅对区间感兴趣，而且通过定义波雷尔 σ-域我们可以确信，在集合运算中事件空间是闭合的(最小的 σ-域包含了$\mathcal{R}$上所有的开区间)。孤立点的概率肯定是 0，任意一个区间可以被赋非零的概率值，并且满足概率论公理。这就引出了第 3 章将要介绍的累积分布函数(cdf)和概率密度函数(pdf)。波雷尔 σ-域是幂集的一个子集，即$\mathcal{B}(\mathcal{R})\subset\mathcal{P}(\mathcal{B})$，并且从基数上比较$\beth_1$ 比$\beth_2$“更加小”。无限可数集合(比如$\mathcal{Z}^+$)与幂集的基数是相同的，见表 2-4。

2.15 $\mathcal{R}$的不可测子集

在前面提及波雷尔 σ-域对于概率空间在实际应用中是足够的，因此我们用$\mathcal{B}(\mathcal{R})$来描述$\mathcal{R}$上的事件。$\mathcal{B}(\mathcal{R})$并非是$\mathcal{R}$的唯一子集，实际上存在着无限多个$\mathcal{R}$的子集，但它们并不一定都是可测的，也并不能都被赋予概率值。区间 $[0,1]$ 上所有的开区间生成了波雷尔 σ-域。对于其中任意的开区间 (a,b)，其 L 测度为 $L(a,b)=b-a$，$b>a$，并且该测度不会大于 1。康托尔集合是区间 $[0,1]$ 上的另外一个子集，它不同于波雷尔 σ-域并且是可测的，

而且 L 测度为 0。一般而言，描述出$\mathcal{R}$上不可测的子集比较困难。我们用一个例题来描述，基于构造法生成维塔利集合，该集合的形式非常复杂，在实际情况下几乎不会遇到。

例 2-54（维塔利集合）再次考虑闭区间 [0,1]，我们将要构造出一个不可测的子集（Billingsley，1986）。◀

为讨论方便起见，我们采用 L 测度 $L(a,b)$。如果 x 和 y 的差是有理数，我们就定义这两个数是等价或相似的，记作：$x\sim y$，如果 $x-y\in\mathcal{Q}$。采用如下的步骤来构造维塔利集合。

- 对于每个 $x\in\mathcal{R}$，定义子集$\{y: x\sim y\}$。这样一个子集称作 x 的等价类（equivalence class）。构造这些类的过程，等效于每个 x 只属于的一个等价类。如果 x 同时属于多个类，那么这些类一定是同一个等价类。所有的有理数$\mathcal{Q}$必然在同一个等价类中。因为所有的 $x\in\mathcal{R}$是可数的，并且这些等价类是不相交的（它们必然是不相交的，否则根据定义两个有重叠部分的等价类必然是同一个等价类），所以这些等价类构成了$\mathcal{R}$的一个划分。
- 对于每一个等价类，在区间 [0,1] 之间选择一个“有代表性”的值。根据选择性公理（axioms of choice）这点是可以做到的。对于类而言，在区间 [0,1] 之间选择的是哪个具体的数值并不重要。比如，对于一个有理数集$\mathcal{Q}$，我们可以选择 1/2（其他一些数值也是可以的，比如 1/2，2/3 等）。
- 用 V 来表示区间 [0,1] 中这些有代表性的数值集，该集合组成了维塔利集合。通过这种方法，我们构建了区间 [0,1] 上一个非常特殊的子集，该集合是不可测的：维塔利集合与波雷尔和康托尔集合不同，它是没有长度的。

表 2-6 列举了集合 V 中的一些元素。每个等价类都是无限可数的，因为它们是根据有理数定义的，而有理数集合是一个可数集合。每个等价类的基数同有理数集$\mathcal{Q}$相同，都是$\beth_0$。但是等价类的数目，以及维塔利集合本身是不可数的。

表 2-6　维塔利集合中的一些元素

等价类的例子	[0,1] 中一些有代表性的数值	等价类的例子	[0,1] 中一些有代表性的数值
有理数集$\mathcal{Q}$	1/2	$\sqrt{2}$，$\sqrt{2}+4$，$\sqrt{2}-20$，…	$\sqrt{2}-1$
π，$\pi-1/2$，$\pi+30$，…	$\pi-3$	e，e−1，e−10，…	e−2

假定 V 是可数的，将其 L 测度记作 $L(V)$。我们通过一个矛盾来证明这个假设是错误的。

- 将原始的维塔利集合进行平移，给出集合定义 $V(x)\triangleq\{v+x:v\in V\},x\in\mathcal{R}$。因为 L 测度是一个平移不变量（见习题 2-19），因此对于任意 x 都有 $L(V(x))=L(V)$。
- 对区间 [−1,1] 中的有理数进行排序，记作$\{q_n\}$，$n\in\mathcal{N}$。构造集合$\{V(q_n)\}$，将 V 平移有理数 q_n 得到 $V(q_n)$。平移量是一个有理数，可以确保平移后得到的仍然是同一种类型的等价类。通过采用对称区间 [−1,1]，原始区间 [0,1] 在两个方向上都得到了拓展，得到的区间 [−1,2] 可用于$\{V(q_n)\}$。
- 集合$\{V(q_n)\}$是不相交的，这点可以通过下面的方法得到验证。设 z 是$\{V(q_m)\}$和$\{V(q_n)\}$的公共元素（$m\neq n$），意思是对于某些 x_m，$x_n\in V$，该公共元素可以写成 $z=x_m+q_m=x_n+q_n$。整理该等式可得 $x_m-x_n=q_n-q_m\in\mathcal{Q}$，这意味着 $x_m\sim x_n$。但是它们不可能属于同一个类，根据 V 的构造可知，V 的每个元素只能来自于不同的类，因此$\{V(q_n)\}$一定是不相交的，得证。
- 这步最关键：因为 $q_n\in[-1,1]$，可得：

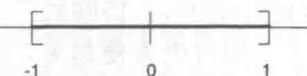
$$[0,1]\subseteq\bigcup_{n=1}^{\infty}V(q_n)\subseteq[-1,2] \tag{2-105}$$

通过上面的方法对维塔利集合进行平移，意味着$\{V(q_n)\}$及其可数并集都不可能超出区间$[-1,2]$（对应于上式第二个$\subseteq$符号），这点在式(2-105)中是很显然的。对于第一个$\subseteq$符号，该平移确保$[0,1]$中所有元素都属于不相交集合$\{V(q_n)\}$的并集。这点可以通过下面的方法来验证。设x是$[0,1]$中任意一个实数，根据维塔利集合的构造过程可知，对于等价类中的元素x，存在一个$v\in V$，使得$x-v$是区间$[-1,1]$中的一个有理数q。对V进行平移，平移量为$[-1,1]$中的有理数，这样可以确保$[0,1]$中的所有数组成公式(2-105)中并集的一个子集。

对于区间[0,1]构造维塔利集合V



在对称区间[-1,1]上对所有的有理数排序



对集合V中的元素进行平移，平移量为[-1,1]上所有的有理数，平移后的元素都包含在扩展区间[-1,2]里面

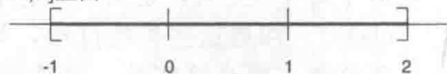


图 2-14 例 2-54 中用到的区间，用于描述维塔利集合

- 图 2-14 总结了各种类型的区间。因为$\{V(q_n)\}$是不相交的，该并集的L测度为：

$$L\Big(\bigcup_{n=1}^{\infty}V(q_n)\Big)=\sum_{n=1}^{\infty}L(V(q_n))\tag{2-106}$$

根据式(2-105)可知：

$$L([0,1])=1\leqslant L\Big(\bigcup_{n=1}^{\infty}V(q_n)\Big)\leqslant 3=L([-1,2])\tag{2-107}$$

在集合V上进行平移是非常有必要的，这样$[0,1]$中所有的实数都被包含在它们的并集里，使得我们可以对可数并集的测度生成一个非零下限$L([0,1])=1$。

- 因为L测度是一个平移不变量，我们期望原始维塔利集合有如下结论：

$$1\leqslant\sum_{n=1}^{\infty}L(V)\leqslant 3\tag{2-108}$$

但是对于常数$L(V)$，对其无限累加求和得到的结果不可能在区间$[1,3]$中，这就是一个矛盾。如果求和区间的序号是n，那么$L(V)$是0同时求和为0，或者$L(V)$是非零同时求和是无穷大。

这个矛盾证明了维塔利集合不可能有长度。通过这种方式构造的集合，使得L测度为平移不变量的事实被推翻了，很显然，这说明最开始的可数假设是不成立的。

在第3章中我们还会讲到连续样本空间，并将定义连续随机变量，包括一些非常著名的参数“家族”，它们在许多实际问题中获得了广泛的应用。

习题

集合运算

2-1 请计算给出样本空间的幂集$\mathcal{P}(\Omega)$：
(a) $\Omega=\{-1,0,1\}$，
(b) $\Omega=\{1,2,3,4\}$，
(c) $\Omega=\{[0,1),[1,2),[2,\infty)\}$。

2-2 请根据维恩图证明$(ABC)^c=A^c\cup B^c\cup C^c$，并画出$((A\cup B)C)^c$。

2-3 已知$\Omega=[-10,10]$，$A=[-1,1]$，$B=[-5,0]$，请计算 (a) AB^c，(b) $A^c\cup B$，(c) $A\oplus B$。

2-4 已知$\Omega=[-10,10]\times[-10,10]$，$A=[0,3]\times[-10,10]$，$B=[-2,1]\times[-10,0]$，请计算 (a) A^cB，(b) AB^c，(c) $B-A$。

2-5 设$\Omega=\mathcal{R}$，定义事件$E_1=\{x: x>6\}$，$E_2=\{x: |x-1|<5\}$，$E_3=\{x: x^2>4\}$。请计算 (a) E_1E_2，(b) $E_1\cup E_2$，(c) $E_2E_3^c$，(d) $E_2\cup E_3$。

2-6 请尝试不使用维恩图证明德摩根定律。(a) $(A\cup B)^c=A^cB^c$，(b) $(AB)^c=A^c\cup B^c$。

2-7 请尝试不使用维恩图证明下面的等式：(a) $A\cup B=(A\oplus B)\oplus(A\cap B)$，(b) $A\cap(B\oplus C)=(A\cap B)\oplus(A\cap C)$。

2-8 笛卡儿乘积适用于有限个元素的集合：如果$A=\{a_1,a_2\}$，$B=\{b_1,b_2\}$，那么$A\times B=\{\{a_1,b_1\},\{a_1,b_2\},\{a_2,b_1\},\{a_2,b_2\}\}$。请证明一般情况下的$\{A,B,C\}$。(a) $A\times(B\cup C)=$

$(A\times B)\cup(A\times C)$，(b) $|A\times B|=|A||B|$。

域

2-9 已知样本空间 $\Omega=\{-2,-1,1,2\}$。(a) 对于事件 $A=\{-1,1\}$ 和 $B=\{-2,2\}$，证明 $\{\phi,\Omega,A,B\}$ 是一个域。(b) 对于事件 $C=\{-2,-1\}$，$D=1$，$E=2$，证明 $\{\phi,\Omega,C,D,E\}$ 不是一个域。(c) 可否通过对 $\{\phi,\Omega,C,D\}$ 添加事件来构建一个非平凡域，并且其基数小于 $|P(\Omega)|=16$。

2-10 设样本空间 $\Omega=[0,10]$。判断下面的集合是否是域？(a) $\{[0,x]\}$ 类型的所有区间的集合。(b) $\{[x,y]\}$ 类型的所有区间的集合。

2-11 设 $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 是样本空间 Ω 的 σ-域，判断 $\mathcal{F}_1\cap\mathcal{F}_2$ 是否是一个 σ-域。

2-12 设 $\Omega=\{-2,-1,0,1,2\}$，对于事件 $A=\{-2,-1,0\}$，$B=\{1,2\}$，$C=\{-2,1,0\}$，$D=\{-1,2\}$，判断下面哪个是域？(a) $\{\phi,\Omega,A,B\}\cap\{\phi,\Omega,C,D\}$，(b) $\{\phi,\Omega,A,B\}\cup\{\phi,\Omega,C,D\}$。

2-13 对于样本空间 $\{\Omega,\mathcal{F}\}$，设 Ω 被划分为了有限个集合 $\{E_1,\cdots,E_N\}$。例如，一种可能的划分为 $\Omega=\mathcal{R}$，$E_1=(-\infty,-1]$，$E_2=(-1,-1]$，$E_3=(1,\infty)$。请证明一般性的结论：$\cup E_n\in\mathcal{F}$，以及 $\cap E_n\in\mathcal{F}$。

2-14 请给出样本空间 $\Omega=[0,10]$ 的最小域，以 $\{\phi,\Omega,0,[0,2)\}$ 开头，并确定这个域的原子。

2-15 已知样本空间 $\Omega=\{1,2,3,4\}$，以 $\{\phi,\Omega,1,\{3,4\}\}$ 开头，再次重复求解习题 2-14。

2-16 请证明自然数集 $\mathcal{N}$ 是一个波雷尔集合，同样的，证明无理数集合 $\mathcal{R}-\mathcal{Q}$ 也是一个波雷尔集合。

测度

2-17 给出下面这些样本空间的计数测度：(a) $\Omega_1=\{1,\cdots,10\}\cup\{1,15,18\}$，(b) $\Omega_2=\{1,2,3\}\times\{2,6\}$，(c) $\Omega_3=Z^+\cap[0,5]$。

2-18 给出下面这些样本空间的 L 测度：(a) $\Omega_1=[0,1]\times[2,5]$，(b) $\Omega_2=[1,2]\times[10,12]\times[2,3]\times[-2,0]$，(c) $\Omega_3=\mathcal{N}\cap[0,10]$。

2-19 设 E 是样本空间 $\Omega=\mathcal{R}$ 的一个事件，请证明对于所有的 $a\in\Omega$，$L(E+a)=L(E)$，也即是说 L 测度是一个平移不变量。

2-20 已知区间集合 $\{|x-n|\leqslant 1/3^n\}$ 定义在 $\mathcal{R}^+$ 上，$n\in\mathcal{N}$，例如当 $n=1$ 时，区间为 $[2/3,4/3]$。请计算这个集合的 L 测度。

2-21 请证明式(2-30)给出的标准高斯测度是一个测度。

2-22 设 E 是 $\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$ 的一个事件，请证明计数测度 $C(E\cap\mathcal{Q})$ 是可数可加的。

2-23 请证明狄拉克测度 $\delta_\omega(E)$ 是一个概率测度，并解释为什么 E 是 Ω 中一个几乎必然事件。

基本概率论

2-24 对于一般事件 A 和 B，请证明：(a) $P(A\cup B)\geqslant 1-P(A^c)-P(B^c)$，(b) $P(A)\leqslant P(B^c)$。

2-25 对于一般事件 E 和 F，请证明：(a) $P(E-F)=P(E)-P(EF)$，(b) $P(E)+P(F)+P(E^cF^c)\geqslant 1$。

2-26 请证明下面的不等式：

$$P(E_1\cup\cdots\cup E_N)\leqslant\sum_{n=1}^{N}P(E_n)\tag{2-109}$$

2-27 (a) 请证明 Bonferroni 不等式：

$$P(E_1\cdots E_N)\geqslant 1-\sum_{n=1}^{N}P(E_n^c)\tag{2-110}$$

(b) 请证明下面的不等式：

$$P(E_1\cdots E_N)\geqslant\sum_{n=1}^{N}P(E_n)-N+1\tag{2-111}$$

2-28 已知事件 $E_1\subset E_2\subset\cdots$，定义 E 是它们的并集。请根据概率论公理证明：

$$\lim_{n\to\infty}P(E_n)=P(E)\tag{2-112}$$

条件概率、全概率和独立性

2-29 对于掷骰子试验，设 $E=5$，$F=\{$至少为 3$\}$，$G=\{$奇数$\}$。请计算下面的概率值：(a) $P(E|F)$，(b) $P(G|F)$，(c) $P(E|FG)$。

2-30 请判断下面命题的真伪。(a) 如果 $P(E|F)=P(E|F^c)$，那么 E 和 F 是相互独立的。(b) 如果 $P(E|F)\geqslant P(E)$，那么 $P(F|E)\geqslant P(F)$。

2-31 假设一个机器由 5 个独立的器件组成，每个器件失效的概率为 0.01。(a) 计算器件都不失效的概率，以及器件全部失效的概率。(b) 计算至少有两个器件同时失效的概率。

2-32 对于投掷两个骰子的试验，设 $E=\{$有一个骰子点数为偶数$\}$，$F=\{$点数之和为 10$\}$，$G=\{$第二个骰子的点数为 4$\}$，请判断这些事件是否是相互独立的。

2-33 设事件 E 表示一个人患某种疾病。当一个病人患这种疾病时，设某医学诊断测试的错误概率为 0.01。设事件 F 表示该测试呈阳性，并假设 $P(F|E^c)=P(F^c|E)=0.05$。如果一个人患这种疾病的概率为 0.01（不进行任何测试的先验知识），请计算 $P(E|F^c)$。

2-34 事件 E 和 F 是相互独立的，请判断：(a) E^c 和 F^c 是否也是相互独立的？(b) E^c 和 F 是否是相互独立的？

2-35 对于事件 E，F 和 G，请证明：(a) $P(E|G)=P(EF|G)+P(EF^c|G)$。(b) 如果 $E\subseteq F$，$P(E|G)\leqslant P(F|G)$。

2-36 对于事件 E，F 和 G，请证明：(a) 如果 $F \subset E$，$P(E|F)=1$。(b) $P(EF|G)=P(E|FG)P(F|G)$。

2-37 请证明或反驳下面的结论：(a) $P(A|B)>P(A) \Rightarrow P(B|A)>P(B)$。(b) $P(A)>P(B) \Rightarrow P(A|C)>P(B|C)$。(c) 如果 $P(A)=1/3$，$P(A \cup B)=1/2$，那么 $1/6 \leqslant P(B) \leqslant 1/2$。

2-38 三个外表完全相同的盒子装有若干彩色球，包括蓝球(B)，红球(R)和黄球(Y)。第一个盒子中有 15 个蓝球，26 个红球，32 个黄球；第二个盒子中有 22 个蓝球，10 个红球，18 个黄球；第三个盒子中有 10 个蓝球，23 个红球，30 个黄球。如果有一个球被挑出，请计算下面的概率：(a) $P(B)$，(b) P(第 1 个盒子$|R$)，(c) P(第 2 个盒子$|B$或Y)。

有限样本空间的概率

2-39 同时投掷四个骰子，试计算下面这些概率：(a) P(四个骰子的点数相同)。(b) P(四个骰子的点数都不同)。(c) P(三个骰子点数相同)。

2-40 在由 N 个人组成的人群中，请计算至少两个人生日是同一天的概率。(假设一年有 365 天，并忽略闰年。)

2-41 某考试中有 5 个选择题，每题有 4 个选项。如果一个学生完全靠猜测来作答，请问：(a) 全部猜错的概率是多少？(b) 猜对三个或三个以上的概率是多少？

2-42 一个钱包里面有 4 张一美元的钞票，6 张十美元的钞票。先取出一张钞票，然后将这张钞票放回钱包，同时还放入 3 张相同面值的钞票，请计算下面这些概率：(a) P(两次共取出 11 美元)，(b) P(三次共取出 30 美元)，(c) P(四次共取出 22 美元)。

2-43 对于一个三位数字 $a_1a_2a_3$，第一位数字的字母表是 $\{0,1\}$，第二位和第三位数字的字母表是 $\{0,\cdots,5\}$。计算下面这些概率：(a) $P(a_1a_2a_3=100)$，(b) $P(a_1a_2a_3>100)$，(c) $P(100<a_1a_2a_3<150)$。

2-44 一个电路包含 8 个电阻，2 个电感和 3 个电容。已知该电路上有三个元器件被损坏了，计算下面这些概率：(a) P(只有一个电容被损坏)，(b) P(至少有一个电感被损坏)。(c) P(没有电阻被损坏)。

2-45 在扑克牌游戏中手里握有 5 张牌，计算下面这些概率：(a) P(5 个红心)(俗称同花)。(b) P(5 个连续牌)(俗称同花顺)，(c) P(5 张牌的点数都不同)。

2-46 一个盒子里面装有若干个外表完全相同的球，只能靠颜色来进行区分，包括 4 个红球，5 个蓝球和 8 个黄球。如果从里面选出三个球(非替代性的)，计算下面的概率值：(a) P(3 个红球)，(b) P(没有黄球)，(c) P(3 个不同颜色的球)。

2-47 同时投掷两枚骰子，计算下面这些概率：(a) P(点数之和大于 8)，(b) P(点数之和位于 4 和 6 之间)，(c) P(两个点数之差为± 2)。

2-48 将两枚骰子投掷 N 次。(a) 以 N 来表示组合$\{1, 2\}$至少出现一次的概率。(b) 当 N 取多少时可使该概率值超过 0.9。

2-49 一个样本空间含有 N 个元素，请证明从中选出 M 个元素(替代性选择)的组合数目为$\binom{N+M-1}{M}$。

仿真作业

2-50 请使用 MATLAB 中的 rand 函数，通过对区间 $[0,1]$ 的划分来模拟连续投掷 100 次骰子的过程，并绘制出该试验的两个不同现实。

2-51 使用 randperm 来产生 $n=5$ 个整数的随机排列。请编写一个程序来统计所有 5! 种排列都被随机生成需要的仿真次数。

进一步阅读

本书中主题的进一步讨论可参考以下文献。集合理论：Peebles(1993). Fiesld：Kingman 和 Taylor(1966)，Wong 和 Hajek(1985)，以及 Gray 和 Davisson(2004)。测量理论：Arthreya 和 Lahiri(2006)，Billingsley(1986)，以及 Kingman 和 Taylor(1966)。概率：Bertsekas 和 Tsitsiklis(2002)，Dougherty(1990)，Golberg(1984)，Miller 和 Freund(1977)，Papoulis(1990)，Peebles(1993)，Ross(1976)，Williams(1991)，以及 Yates 和 Goodman(2005)。

第3章 随机变量

3.1 引言

第2章介绍了测度理论，还将σ-域$\mathcal{F}$中事件E的概率定义为了一种测度，该测度将区间$[0,1]$中的一个数字赋给该事件。很容易从数学上将$\mathcal{F}$中的抽象事件映射到实线$\mathcal{R}$上。在大多数应用中可以直接在$\mathcal{R}$上讨论事件，而不必理会底层的抽象样本空间Ω。在许多工程问题中需要用到$\mathcal{R}$上定义的随机变量，一个原因是工程问题往往通过数值来进行定义和评估(这与Ω中的抽象量完全不同)，另外一个原因是因为区间在连续随机试验中非常重要，这在第2章中已经得到了讨论。对于连续随机试验，不可能给独立点赋非零的概率值，而非零概率的事件是通过$\mathcal{R}$上的区间来进行定义的，这样才能满足概率论三大公理。波雷尔σ-域$\mathcal{B}(\mathcal{R})$是最小的σ-域，它由$\mathcal{R}$上所有的开区间生成，并且所有的波雷尔集合可以以一致的方式被赋予概率值。

本章的一个目的是定义从抽象样本空间Ω到实线$\mathcal{R}$的一个映射，并以此得出随机变量X。另外一个目的是描述给$\mathcal{B}(\mathcal{R})$中事件的各种概率赋值方法，即各种实际应用中的模型。所有的概率都可以从累积分布函数(cdf)中推导出来，定义为$F_X(x)\triangleq P(X\leqslant x)$，表示$X$位于半开半闭区间$(-\infty, x]$内的概率。根据概率论三大公理，可知$0\leqslant F_X(x)\leqslant 1$，当$x$在$\mathcal{R}$上进行变化时，根据$F_X(x)$的表达式或形状就可定义出某种特定的随机变量。图3-1给出了离散和连续随机变量的例子。离散随机变量的累积分布函数必然是一个“阶梯”函数，而连续随机变量的累积分布函数是连续且平滑的。(稍后给出图3-1a中圆圈的定义。)两种类型的累积分布函数都是非递减的。在本章中，我们将会看到许多类型的$F_X(x)$，它们都很重要，因为在实际问题中为了给事件的“可能性”建模，将会用到某个特殊的概率赋值方法。

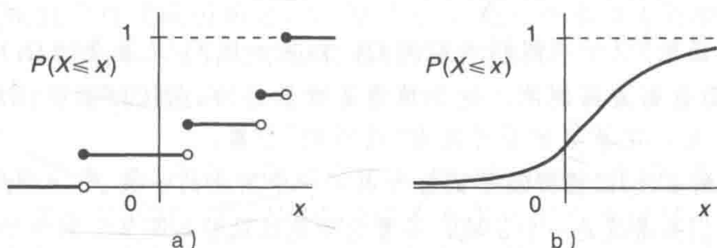


图3-1 累积分布函数$F_X(x)$的一些例子。a) 离散随机变量；b) 连续随机变量

3.2 函数和映射

之前概率测度被用到了事件空间$\{\Omega, \mathcal{F}\}$上，它满足三大特性，并且可以提供这些事件的容量信息。这些特性被称作概率论三大公理。维塔利集合是$\mathcal{R}$的一个不可测子集的例子。另一方面，波雷尔σ-域是可测的，并且很容易用L测度来进行描述。

例3-1 (波雷尔集合) 设$[a,b]$是$\mathcal{B}(\mathcal{R})$中的一个闭区间，$a\leqslant b\in\mathcal{R}$。可知：(ⅰ)$L(a, b)=b-a\geqslant 0$，(ⅱ)$L(\phi)=0$，(ⅲ)如果$[c,d]$是另外一个区间，$c\leqslant d\in\mathcal{R}$，并且满足$[a,b]\cap[c,d]=\phi$，那么$L([a,b]\cup[c,d])=L([a,b])+L([c,d])=(b-a)+(d-c)$。 ◀

定义(可测集合) **可测集合**是指 Ω 中元素的集合，并且满足测度的三大性质。如果 $\mathcal{F}$ 是事件的一个可测集合，那么 $\{\Omega,\mathcal{F}\}$ 是一个可测的事件空间。

在定义可测函数之前，我们再次研究抽象样本空间 Ω。对于有限样本空间，我们可以设想 Ω 的一个简单例子，它的元素不直接来自于实线。这包括投掷一枚硬币或一对骰子，以及其他很多方法。对于无限可数试验和不可数试验，类似的非数值例子不太容易想象。对于连续样本空间，Ω 可以是所有正弦波形 $A\sin(2\pi ft)$ 的集合，$t\in[0,T]$，$A\in[0,1]$。因此，Ω 包含了无限多个正弦，这些正弦的时延都是 T，频率都是 f，但幅度是连续的。当然，这个试验与数值样本空间 $\Omega=[0,1]$ 其实是等效的。即使我们采用 $\{\Omega,\mathcal{F}\}$ 来表示一个抽象的事件空间，目的就是为了定义一个叫做随机变量的可测函数，而在实际情况下"原始的"样本空间往往是 $\Omega=\mathcal{R}$(也就是说，在特定的应用场合，基础绝对样本空间有可能不存在)。

考虑可测事件空间 $\{\Omega_1,\mathcal{F}_1\}$，其中 $\mathcal{F}_1$ 是样本空间 Ω_1 的一个 σ 域。设 $\{\Omega_2,\mathcal{F}_2\}$ 是另外一个可测事件空间，$\mathcal{F}_2$ 是样本空间 Ω_2 的一个 σ-域。我们希望给出从 Ω_1 的输出到 Ω_2 的一个映射，并且第一个事件空间中的测度可以搬移到第二个事件空间中去。

定义(可测函数) 函数 $g(\omega)$ 的定义域为 Ω_1，值域为 Ω_2，如果对于任意事件 $A\in\mathcal{F}_2$，下面的逆象成立，则 $g(\omega)$ 是一个**可测函数**:

$$g^{-1}(A)=\{\omega:g(\omega)\in A\}\in\mathcal{F}_1 \tag{3-1}$$

$\mathcal{F}_2$ 中事件的逆象必须通过一个可测函数被包含在原始的 σ-域 $\mathcal{F}_1$ 里面。根据定义，一个函数总可以将 $\omega_1\in\Omega_1$ 唯一映射到 $\omega_2\in\Omega_2$，但对于逆象该结论不成立。图 3-2 给出了一个可测函数的直观表示，可看出 $\mathcal{F}_2$ 中事件的逆象位于 F_1 里面。因为我们的目的是要给 $\mathcal{F}_1$ 中的事件赋予概率值，也就是将 $\mathcal{F}_2$ 中对应事件的概率值进行映射，此时只有关于 $\mathcal{F}_1$ 的逆象是闭合的才能保证满足概率论三大公理。

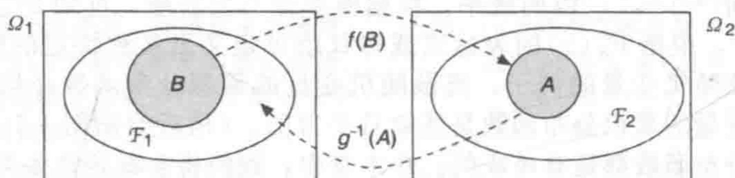


图 3-2 两个样本空间之间一个可测函数的直观表现，$B\in\mathcal{F}_1$ 的输出映射到 $A\in\mathcal{F}_2$ 的元素，A 的逆象的元素并不都属于 B，但仍然属于 $\mathcal{F}_1$

例 3-2 (可测函数)对于有限样本空间 Ω，如果 σ-域 $\mathcal{F}_1$ 是幂集 $\mathcal{P}(\Omega_1)$，那么从 Ω_1 映射到 Ω_2 的任意一个函数都是可测的。这个结论是很显然的，因为 $\mathcal{F}_2$ 中任意一个事件的逆象都属于 $\mathcal{F}_1$，根据定义，在幂集中不可能有"遗漏的"子集。 ◀

例 3-3 (不可测函数)以投掷骰子试验为例，一个简化的 σ 域 $\mathcal{F}_1=\{\phi,\Omega_1,\{1,3,5\},\{2,4,6\}\}$。在例 2-23 中我们知道这是一个 σ 域，尽管它不是该试验的幂集。对于如下从 Ω_1 到 Ω_2 的平凡映射:

$$g(\omega)=\omega,\quad \omega=1,2,3,4,5,6 \tag{3-2}$$

设 $\mathcal{F}_2$ 是 Ω_2 中元素的幂集，并且 $|\mathcal{F}_2|=2^6=64$。很显然，$\mathcal{F}_2$ 中部分事件的逆象不可能是 $\mathcal{F}_1$ 中的事件。比如，3 的逆象就不是 $\mathcal{F}_1$ 中的事件，尽管 3 是该试验的一个输出结果，这个特殊的 $\mathcal{F}_1$ 把独立的输出结果排除在外。通过选择 $\mathcal{F}_1$ 为幂集的一个子集，许多潜在的事件被排除在该试验之外。 ◀

对于一个可测函数，需要注意下面几点:

- 函数 $g(\omega)$ 是对样本空间 Ω_1 中所有的元素(输出)进行定义的，并不是对 F_1 中的事件。
- 一个函数的可测性取决于两个样本空间中特定的 $\mathcal{F}_1$ 和 $\mathcal{F}_2$。如果其中任意一个 σ-域被修改了，也可能会导致函数的可测性发生变化。

对于上面最后一点，假设例 3-3 中 $\mathcal{F}_2=\{\phi,\Omega_2\}$ 为一个平凡的 σ-域。此时，$g(\omega)$是一个可测函数。实际上，从 Ω_1 到 Ω_2 的任意一个映射都可能是可测的，因为对于$\mathcal{F}_2$ 中的两个事件，有下面的结论：

$$g^{-1}(\phi)=(\omega:g(\omega)=\phi)=\phi\in\mathcal{F}_1 \tag{3-3}$$

$$g^{-1}(\Omega_2)=\{\omega:g(\omega)=\Omega_2)=(1,2,3,4,5,6)=\Omega_1\in\mathcal{F}_1 \tag{3-4}$$

例 3-4 (可测函数)在例 3-3 中，很显然存在无数个不可测的映射。下面的函数是可测的，如图 3-3 所示。

$$g(\omega)=\begin{cases}0, & \omega=1,3,5\\ 1, & \omega=2,4,6\end{cases} \tag{3-5}$$

这个函数将原始的样本空间压缩为了一个 $\Omega_2=\{0,1\}$ 的二元试验：我们只关心骰子的点数是奇数还是偶数。假设该 σ-域的幂集使得 $\mathcal{F}_2=\{\phi,\Omega_2,0,1\}$，可以看出这个映射是可测的：

$$g^{-1}(0)=\{\omega:g(\omega)=0\}=(1,3,5)\in\mathcal{F}_1 \tag{3-6}$$

$$g^{-1}(1)=\{\omega:g(\omega)=1\}=\{2,4,6\}\in\mathcal{F}_1 \tag{3-7}$$

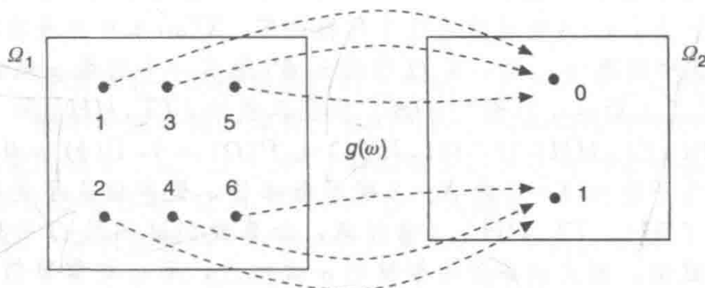


图 3-3 例 3-4 中的映射

从上面这些有限样本空集的例子可知：

- 如果有输出(基本事件)被排除在$\mathcal{F}_1$ 之外，此时很容易找出不可测的函数，尤其当$\mathcal{F}_2$ 为幂集的情况。
- 如果 Ω_1 被压缩为一个元素数目更少的 Ω_2，此时的函数往往是可测的。

在例 3-4 中上面这个性质就很明显，在那里 $g(\omega)$是一个多对一的映射。对于无限可数以及不可数(连续)样本空间，我们期望也有类似的结论成立。◀

定义(随机变量) **随机变量** X 是 Ω 到实线$\mathcal{R}$的一个可测函数。对于 $\omega\in\Omega$，该函数记作 $X(\omega)$。

抽象样本空间被记作 $\{\Omega,\mathcal{F},P\}$。随机变量 X 对应的概率空间是$\{\mathcal{R},\mathcal{B}(\mathcal{R}),P_X\}$，其中$\mathcal{B}(\mathcal{R})$是实线上的波雷尔 σ-域，P_X 是$\mathcal{B}(\mathcal{R})$中所有事件的概率测度。因为 $X(\omega)$是可测的，所以$\mathcal{B}(\mathcal{R})$中任意一个事件一定存在一个位于$\mathcal{F}$中的逆象。也就是说，

$$X^{-1}((-\infty,x])=\{\omega:X(\omega)\leqslant x\}\in\mathcal{F} \tag{3-8}$$

请注意，在第 1 章中，我们也用 $X(\omega)$来表示 $x(t)$的傅里叶变换。尽管从上下文很容易区分记号的具体含义，但我们很快就会在映射中省去 $X(\omega)$的自变量，改为采用 X 来表示一个随机变量。

从前面的讨论可知，一个函数仅凭自身是不可测的：它的可测性需取决于$\mathcal{F}$，在前面我们并没有特意强调这点。因此，如果我们说 $X(\omega)$是可测的，或者称其为一个随机变量，那么我们假设在抽象事件空间中采用了一个合适的$\mathcal{F}$，比如 Ω 的幂集。

例 3-5 设 $\Omega=\{-3,-2,-1,1,2,3\}$，函数 $X(\omega)=|\omega|$。对于随机变量 X，我们希望找到一个 σ-域$\mathcal{F}$而不是幂集$\mathcal{P}(\Omega)$。通过观察可知逆象为 $X^{-1}(1)=\{-1,1\}$，$X^{-1}(2)=\{-2,2\}$，$X^{-1}(3)=\{-3,3\}$。可以根据这些原子生成 Ω 的 σ-域(见第 2 章)。将 σ-域定义

为原子的幂集，可得 $F=\{\phi,\Omega,\{-1,1\},\{-2,2\},\{-3,3\},\{-3,-1,1,3\},\{-3,-2,2,3\},\{-2,-1,1,2\}\}$。这个 σ-域含有 8 个元素，而 $\mathcal{P}(\Omega)$ 有 $2^6=64$ 个。$X(\omega)$ 为事件空间 $\{\Omega,\mathcal{F}\}$ 的随机变量。◀

例 3-6 假设在例 3-5 中，我们扩展样本空间到除 0 以外的所有整数：$\Omega=\{\cdots,-3,-2,1,1,2,3,\cdots\}$。对于函数 $X(\omega)=|\omega|$，其逆象为 $X^{-1}(n)=\{-n,n\},n\in\mathcal{N}$。如果 Ω 的 σ-域 $\mathcal{F}$ 由原子 $\{-n,n\}$ 产生，那么 $X(\omega)$ 为一个随机变量，它也是 $\mathcal{F}$ 上的一个可测函数。请注意，这些原子的幂集比 Ω 的幂集小。$\{-1,3\}$ 和 $\{1,2,3\}$ 是 $\mathcal{P}(\Omega)$ 中事件的例子，但它们不在原子 $\{-n,\ n\}$ 的幂集中。◀

例 3-7 投掷两枚硬币试验的样本空间为 $\Omega=\{HH,TT,HT,TH\}$，定义如下映射：$X(\omega)=\{$正面朝上的次数$\}$。因此有 $\{TT\}\to 0$，$\{TH,HT\}\to 1$，$\{HH\}\to 2$。如果 $\mathcal{F}=\mathcal{P}(\Omega)$，那么 $X(\omega)$ 是可测的并且是一个随机变量。假设硬币试验是客观公正的，我们可知 $P(X=0)=1/4=P(X=2)$，$P(X=1)=1/2$。假设 $\mathcal{F}$ 不再是幂集，此时 $\mathcal{F}=\{\phi,\Omega,\{TT,HH\},\{TH,HT\}\}$，且 $\mathcal{F}$ 为一个合法的 σ-域。此时，$X(\omega)$ 是不可测的，即使仍然采用上面的映射。因为在这种情况下，逆象 $X^{-1}(2)=\{HH\}$ 并不在这个“粗糙”的 $\mathcal{F}$ 中。通过这个例子，我们强调以下几点：(ⅰ)通过使用这个粗糙的 $\mathcal{F}$，$X(\omega)$ 本身并没有变化。(ⅱ)概率空间中的概率赋给了 $\mathcal{F}$ 中的事件，而不是 Ω 中的元素(除非一个可数 σ-域的原子正好是该样本空间的输出)。对于上面这个粗糙的 σ-域，我们只能给 $\{TT,HH\}$ 和 $\{TH,HT\}$ 赋予概率值，同时可知，$P(\{TT,HH\}\cup\{TH,HT\})=P(\Omega)=1$，$P(\phi)=0$。之所以说 $X(\omega)$ 是不可测的，这是因为不能给 $X=0$ 或 $X=2$ 赋予概率值，概率值是不能赋予独立输出结果的，即使它们组成了原子 $\{TT,HH\}$。(请注意：如果我们试图给 Ω 的输出结果赋予概率值，而不是给事件赋值，而这些事件由粗糙的 σ-域描述，那么就需要假定 $\mathcal{F}$ 的幂集。)对于粗糙的 σ-域，一个可测函数为：

$$X(\omega)=\begin{cases}0, & \text{输出不同}\\ 1, & \text{输出相同}\end{cases} \tag{3-9}$$

所以有 $\{HT,\ TH\}\to 0$，$\{HH,\ TT\}\to 1$。粗糙 σ-域将在第 5 章中进一步介绍。◀

尽管 $X(\omega)$ 的样本空间是实线，但并不是所有的点都需要对应于 Ω 中元素的映射(在前面的例子是很显然的)。一般情况下我们认为：如果 Ω 是一个有限(离散)样本空间，那么在 $\mathcal{R}$ 上只有有限个点的概率值是非零的。每个点都拥有一个概率质量，该质量是 $[0,1]$ 之间的某个数，而且所有点的概率质量之和一定是 1。概率质量为 0 的点绝对不会出现：它们是不可能事件，而且在原始样本空间中没有对应的事件用于映射(否则它们就应该拥有非零的概率值)。$\mathcal{R}$ 中离散点的概率分配称为概率质量函数(pmf)。

例 3-8 对于硬币投掷试验，$\Omega=\{H,T\}$，$\mathcal{F}=\{\phi,\Omega,H,T\}$。设随机变量 $X(\omega)$ 的映射为 $H\to 1$，$T\to -1$。很显然，对于一个公平的试验 $P(X=-1)=P(X=1)=1/2$，X 取其他值的概率都为 0。也就是说，其他数字是不可能事件。如果假设存在硬币立起来这种事件，用 E 来表示，此时的映射为：$E\to 0$。如果 $P(E)\neq 0$(尽管发生这种情况的次数微乎其微)，那么这就是一个完全不同的试验：存在着三个非零概率的输出。而除此之外 X 取其他数字的概率仍然为 0，仍然是不可能事件。◀

通过例 3-8 可知，对于有限试验，零概率事件同时也是不可能事件。类似的结论可适用于无限可数的样本空间：映射到实线 $\mathcal{R}$ 上得到可数个点，每个点的概率质量都不是零。我们的直觉认为，无限多个点的概率质量之和竟然是一个有限值，这似乎是不可能的事。但是，这种情况非常类似于一个收敛的无限长序列(见附录 E)。因为有无限多个样本点，只有特定的概率质量函数之和为 1。后面将通过专门的例子进行讲解。

连续样本空间的情况有很大的区别。因为 Ω 中有无限不可数个元素要映射到实线上，所以对于 $\mathcal{R}$ 上的任意一个独立点其概率质量都不可能是非零的。对于连续随机变量而言，

实线上任意一个点的概率都是0。但是，在第2章曾经提到，概率为0并不意味着该事件是不可能事件。实际上，对于一个连续试验，无限多的零概率点都是有可能出现的。不可能的输出意味着它在实线$\mathcal{R}$上对应的点被排除在了样本空间的定义之外。例如，如果令$\Omega=\mathcal{R}^+$，就使得所有的负数都成为了不可能事件，反之，尽管所有非负数的概率都是0，但它们都是有可能发生的。

因为我们用波雷尔σ-域来描述实线上的事件，所以$\mathcal{R}$上的区间被赋予了非零概率值，这些区间的L测度都是非零的。区间$[a,b]$，$[a,b)$，$(a,b]$和(a,b)被赋予的概率值都是相同的，因为独立点的概率值都是0。接下来将会给出概率密度函数(pdf)的定义，并用它来描述$\mathcal{R}$上的概率测度。一段特定区间$[a,b]$的概率就是pdf下方在a和b之间的面积。严格来说，这种概率是根据L测度定义的。

最后再提及一下混合随机变量，这种变量包括连续和离散随机变量成分。对于这种变量，$\mathcal{R}$上的一些点用概率质量函数来描述，而一些区间又用概率密度函数来描述。不过，所有概率质量函数之和，以及概率密度函数下方面积之和，一定为1(这样才能满足概率论三大公理)。混合随机变量的例子在后面会给出。

3.3 分布函数

为了描述连续、离散和混合随机变量，首先介绍一下分布函数，通过分布函数才能推导出概率质量函数、概率密度函数、或者二者的组合。当随机变量的类型被确定后，往往没有必要去考虑抽象的概率空间$\{\Omega,\mathcal{F},P\}$(前面提到过，这种抽象的概率空间还有可能不存在)。我们直接研究概率空间$\{\mathcal{R},\mathcal{B}(\mathcal{R}),P_X\}$的分布函数。

定义(分布函数) **分布函数**的定义域为$\mathcal{R}$，表达式为下面的概率：

$$F_X(x)\triangleq P_X((-\infty,x])=p(X\leqslant x) \tag{3-10}$$

其中$P((-\infty,x])$是事件空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$内从$(-\infty,x]$到$[0,1]$中一个实数的映射。

这种概率的赋值方法是另外一种类型的映射：从$\mathcal{R}$上一个半开半闭的区间映射到闭区间$[0,1]$上的一个点。图3-4给出了绝对样本空间Ω中的事件E如何映射到$\mathcal{R}$上的一段区间，然后再到闭区间$[0,1]$上的一个点的过程，这个点的数值就是该区间的概率值(也是基础事件E的概率值)。

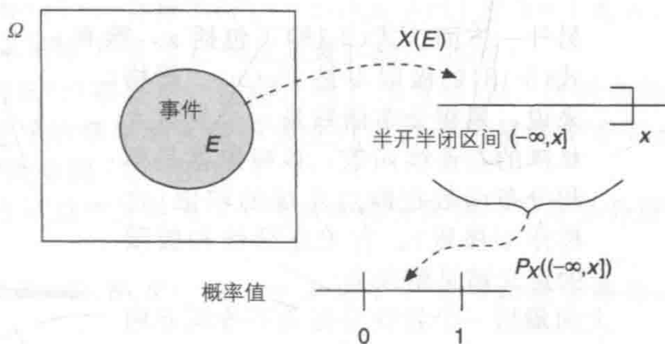


图3-4 从$E\in\Omega$到$\mathcal{R}$上一段区间的映射，然后再到闭区间$[0,1]$上的一个点

关于符号，大写字母(比如X)表示随机变量，而小写的x表示该随机变量的一个特定取值(输出)。$F_X(x)$中的下标是用来强调我们指的是随机变量X，在后面的多维随机变量的分布函数中这个下标就非常有必要。但是，如果上下文的所指非常明显，为行文简洁我们往往会省去这个下标。

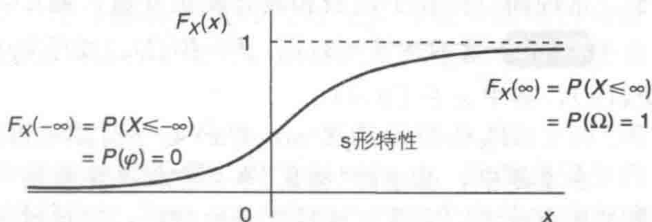


图3-5 累积分布函数$F_X(x)=P(X\leqslant x)$的一个例子

分布函数也称作概率分布函数，以及累积概率分布函数。为了不与后面的概率密度函数混淆，我们用$F_X(x)$来表示累积分布函数(cdf)。可以用抽象概率空间$\{\Omega,\mathcal{F},P\}$来表示累积分布函数：

$$F_X(x)=P(\{\omega:X(\omega)\leqslant x\}) \tag{3-11}$$

式(3-11)表示的是不超过数值x的概率，而这些概率是指$\omega\in\Omega$通过函数$X(\omega)$映射到$\mathcal{R}$上

得到的概率值。

图 3-5 给出了累积分布函数的一个例子，我们看到的函数图形是 s 形的，这是因为该函数是非递减的。绝大多数的“标准”概率密度函数都会有这种形状的累积分布函数，原因就在于累积分布函数实际上是概率密度函数下方的面积。但是，也存在非 s 形的例外，比如均匀分布的累积分布函数就是一条直线，但也可以把它看做是 s 形的一种极端情况。

根据累积分布函数的定义，我们给出该函数的一些性质：

- 端点：

$$F_X(-\infty)=0 \quad 和 \quad F_X(\infty)=1 \tag{3-12}$$

根据 $F_X(x)$的定义，代入端点可得：

$$F_X(-\infty)=P(X\leqslant-\infty)=P(\phi)=0 \tag{3-13}$$

$$F_X(\infty)=P(X\leqslant\infty)=P(\Omega)=1 \tag{3-14}$$

注意，$X\leqslant\infty$包含了$\mathcal{F}$中所有的事件，而 $X\leqslant-\infty$则没有包含$\mathcal{F}$中的任何事件，表示一个空集 ϕ。

- 非递减函数：

$$F_X(x_1)\leqslant F_X(x_2) \quad (对于\ x_1\leqslant x_2) \tag{3-15}$$

因为 $P(x_1<X\leqslant x_2)\geqslant 0$，所以对于互斥事件 $A=(-\infty,x_1]$ 和 $B=(x_1,x_2]$ 有：

$$\begin{aligned}F_X(x_2)&=P(X\in A\cup X\in B)\\&=P(X\in A)+P(X\in B)\\&=F_X(x_1)+P(x_1<X\leqslant x_2)\geqslant F_X(x_1)\end{aligned} \tag{3-16}$$

- 右连续：

如果 $F_X(x)$在 x 处有一个断点，那么任取一个 $\varepsilon>0$：

$$F_X(x)=\lim_{\varepsilon\to 0}F_X(x+\varepsilon)=\lim_{\varepsilon\to 0}P(X\leqslant x+\varepsilon) \tag{3-17}$$

根据不等式的方向，$P(X\leqslant x+\varepsilon)$总是包括点 x，因此 $F_X(x)$是右连续的，

$$\lim_{\varepsilon\to 0}F_X(x-\varepsilon)=\lim_{\varepsilon\to 0}P(X\leqslant x-\varepsilon) \tag{3-18}$$

另外一方面，式(3-18)不包括 x，除非 $\varepsilon=0$。因为函数在 $\varepsilon=0$ 有一个断点(阶跃)，式(3-18)的极限不是 $F_X(x)$。严格来说，累积分布函数是一个具有左极限的右连续函数，该极限就是累积分布函数在断点左端的取值(也称作下极限)。有关连续性和极限的定义请见附录 B。

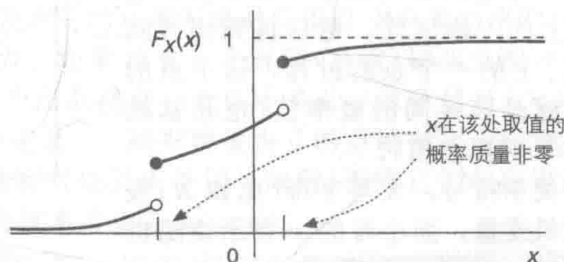


图 3-6 有断点的 $F_X(x)$。实心圆圈表示函数右连续的，空心圆圈表示函数是左断开的

上面最后一个特性是有关不等式方向的，即 $F_X(x)$“小于等于”还是“小于”。图 3-6 给出了一个含有断点的 $F_X(x)$的例子。这种断点存在于离散和混合随机变量，而$\mathcal{R}$中独立点的概率质量为零。

例 3-9 设 $\Omega=[-2,1]$，$\mathcal{F}=\mathcal{B}(\Omega)$。对于映射 $X(\omega)=|\omega|$，我们试着计算分布函数 $F_X(x)$，其中 $x\in[0,2]$：

$$F_X(x)=P(X(\omega)\leqslant x)=p(|\omega|\leqslant x)=P(-x\leqslant\omega\leqslant\min(x,1)) \tag{3-19}$$

在本题中，基础的“抽象”事件空间是实线的一个子集，其 σ-域由区间 $[-2,1]$ 中的波雷尔集合给出。它具有随机变量的特征，可通过另外一个抽象事件空间的映射来产生。假设$\{\Omega,\mathcal{F}\}$的概率测度是 L 测度(乘以 1/3)，其定义见第 2 章。因此可得：

$$F_X(x)=(1/3)L(-x,\min(x,1))=(1/3)[\min(x,1)+x]I_{[0,2]}(x)+I_{(2,\infty)}(x) \tag{3-20}$$

其中 $I_{[a,b]}(x)$是一个指标函数，当 $x>2$ 时，等式只有第二项，此时 $F_X(x)=1$。这个分布

函数是分段线性的：在[0,1]区间的斜率为2/3，在(1,2]区间的斜率为1/3，在其他区间是一条水平线。◀

例3-9的概率空间是$\{\Omega=[-2,1],\mathcal{F}=\mathcal{B}([-2,1]),P(\cdot)=(1/3)L(\cdot)\}$，因为$P$是$L$测度，所以输出$\{\omega\}$服从均匀分布。随机变量$X(\omega)$是从$[-2,1]$到$[0,2]$的一个映射，映射前后都是在实线上，所以实际上这是随机变量变换的一个例子。如果设Y是Ω的一个随机变量，那么Y的分布函数为：

$$F_Y(y)=P(\omega\leqslant y)=(1/3)(y+2)I_{[-2,1]}(y)+I_{(1,\infty)}(y) \tag{3-21}$$

第4章会介绍随机变量的各种变换。本章后续的部分将研究随机变量的特定"系列"，而不去考虑基础的抽象概念空间，并用符号X来替代$X(\omega)$。同时这意味着X是可测的，并且也有明确定义的概率空间$\{\mathcal{R},\mathcal{B}(\mathcal{R}),F_X(x)\}$，其中累积分布函数代表了概率测度$P_X$。

现在，我们对定义在实线上的概率空间有完备的描述，通过累积分布函数也可以对半开半闭区间$(-\infty,x]$进行描述。在接下来的讨论中，非常有必要对离散随机变量、连续随机变量，以及混合随机变量进行区分。现在回顾一下，σ-域是给样本空间定义的，于是我们能够以统一的方式给感兴趣的事件赋概率值。波雷尔σ-域不仅是可测的，而且在实际情况下对于$\mathcal{R}$上的连续随机变量非常有用，我们往往关注的是区间的概率值，比如$P(a\leqslant X\leqslant b)$。

为了区分离散随机变量和连续随机变量，定义一个连续的概率测度。

定义(绝对连续概率测度) 如果对于事件$E\subset\Omega$(区间或点)，$L(E)=0$意味着$P(E)=0$，那么在可测事件空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$上，概率测度$P$关于$L$测度是**绝对连续**的。

其中"关于L测度"指我们感兴趣的是用区间的长度定义测度：如果这段区间的长度为0，那么这段区间的概率为0。一个绝对连续的概率测度引出的是一个连续的随机变量。(在第4章中，该定义被扩展到多维可测事件空间$\{\mathcal{R}^N,\mathcal{B}(\mathcal{R}^N)\}$，其中事件$E$是$\mathcal{R}^2$上的一个矩形，当$N>2$时，$E$是$\mathcal{R}^N$上的一个超矩形。)

我们同样也给出非绝对连续的概率测度的定义。

定义(奇异概率测度) 如果存在事件$\{E_n\}$，使得$L(E_n)=0$以及$P(\bigcup_n E_n)=1$成立，那么在可测事件空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$上，概率测度$P$关于$L$测度是**奇异的**(singular)。

根据定义可知，在$\mathcal{R}$上存在若干点的集合(很显然，这些点的L测度是0)，这些点的并集的概率为1。对于离散随机变量，奇异的概率测度是指独立点的概率质量是非零的。(这个定义同样也可以扩展到多维可测事件空间$\{\mathcal{R}^N,\mathcal{B}(\mathcal{R}^N)\}$。)

勒贝格分解定理指出，一个概率测度可以被分解为一个绝对连续分量和一个奇异分量(Billingsley，1986)，在此不给出证明。

定理3-1 (勒贝格分解) 在事件空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$上，概率测度$P$是可以被分解的，因此对于任意一个事件$E$。

$$P(E)=P_{ac}(E)+P_s(E) \tag{3-22}$$

其中P_{ac}和P_s分别表示的是绝对连续概率测度和奇异概率测度。

这个定理肯定了混合随机变量的存在，混合随机变量可同时拥有离散和连续的分量(在本章最后会讲到)。式(3-22)对每种情况分别进行了处理，连续随机变量的概率测度为P_{ac}，而离散随机变量的概率测度为P_s。

为了内容的完备性，奇异概率测度还可以被进一步分解，一部分包括若干独立点(离散的)，其他部分为无限不可数事件的一个奇异连续分量。在实际问题中，后面这部分内容其实并不很受关注，因此我们并不对其进行详细讲解。接下来给出的这个康托尔分布的例子(Richardson，2009)是一个非绝对连续的随机变量，它实际上是奇异连续的，并且没有离散的部分。

例3-10 (康托尔函数) 康托尔函数$F_C(x)$是指定义在康托尔集合C上的均匀分布随机变量的累积分布函数，在第2章的例2-35中曾对其进行了解释。我们用序列来表示函

数可得到康托尔函数，记作 $F_{C_m}(x)$，$m\in\mathcal{N}$。图 3-7 给出了 m 取 3 个值时的 $F_{C_m}(x)$，可以看出函数曲线呈阶梯状，且当 m 变大时平坦部分越来越多。(这就是所谓的"魔鬼的阶梯"。)当 $m\to\infty$ 时得到的就是康托尔函数 $F_C(x)$，它具有如下特性：(ⅰ)它是一个奇异函数，并且没有离散部分。(ⅱ)它是一个连续的单调函数(很显然，所有的累积分布函数都是单调的)。(ⅲ)它的导数几乎处处为 0，因为对于 $x\notin C_m$，每个 $F_{C_m}(x)$都是常数(每个函数在开区间都是平坦的，并且要被抠去一部分来得到康托尔集合的 C_m)。◀

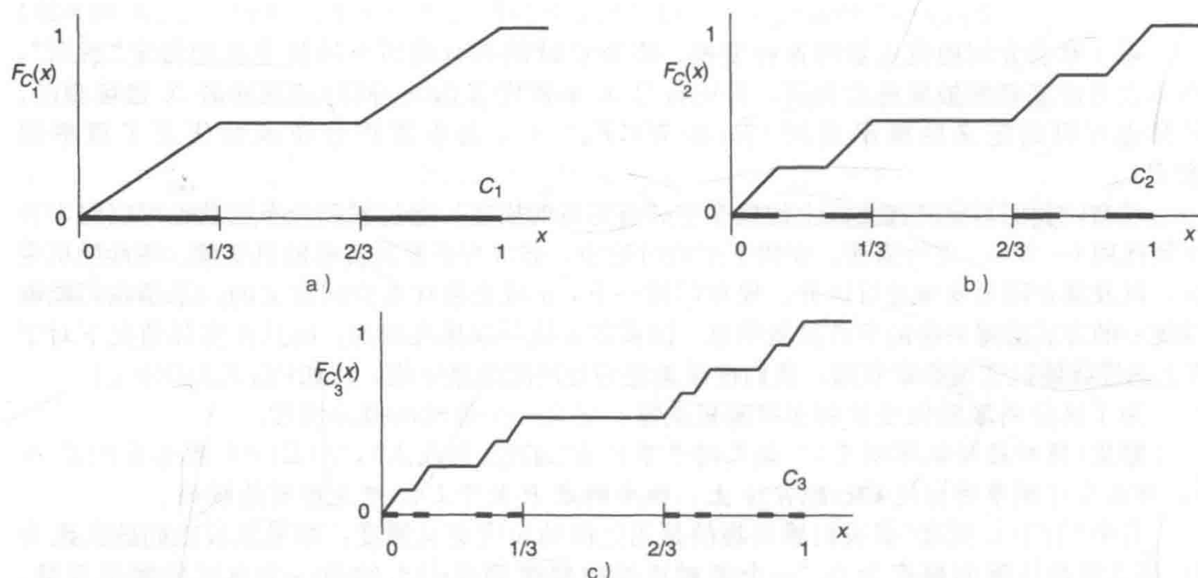


图 3-7　C_m 上均匀分布的累积分布函数，$m=1,2,3$。当 $m\to\infty$时，极限为康托尔函数 $F_C(x)$。从图 2-6 中的$\{C_m\}$得到的康托尔集合也在本图中(x 轴上的粗线)

概率空间 $\{\mathcal{R},\mathcal{B}(\mathcal{R}),F_X\}$ 是对离散、连续和混合随机变量一个非常有用的描述方法。在 3.4 节，对于离散随机变量，我们采用这种描述以及累积分布函数 $F_X(x)$来定量分析独立点的概率质量；在 3.5 节，对于连续随机变量，我们采用同样的方法计算实线上任意一段区间的概率值。离散随机变量的概率质量函数相对容易理解，因此我们首先讲解这个内容。对于连续随机变量，我们利用一段区间来描述一个事件，然后对基础的概率密度函数进行积分来计算该事件的概率值。

本节最后将基于累积分布函数，给出随机变量支集的严格定义。

定义(随机变量的支集)　随机变量 X 的支集由所有的 $x\in\mathcal{R}$给出，且对于所有的 $\varepsilon>0$ 有：

$$F_X(x+\varepsilon)>F_X(x-\varepsilon) \tag{3-23}$$

请注意在这个定义中需要用到严格的不等式。尽管 $F_X(x)$支撑了$\mathcal{R}$，但对于基础的概率密度函数或者概率质量函数，该结论不一定成立(尤其是用到了克罗内克 δ 函数)。开区间被排除在概率密度函数和概率质量函数的支集之外，在开区间上 $F_X(x)$是平坦的(导数为 0)。有平坦区域的$\mathcal{F}_X(x)$或者对应于一个离散随机变量，或者对应于一个具有离散分量的混合随机变量。因为累积分布函数是右连续的，所以上面的这个不等式"捕捉到了"离散随机变量的输出结果。

3.4　概率质量函数

图 3-8 给出了离散随机变量分布函数的一个例子。通过观察可以看出，这种类型的分布函数有两个非常明显的特征：

- 当 x 是随机变量的输出时，$F_X(x)$在 x 处有个突变。

- 在输出值之间，$F_X(x)$为常数。

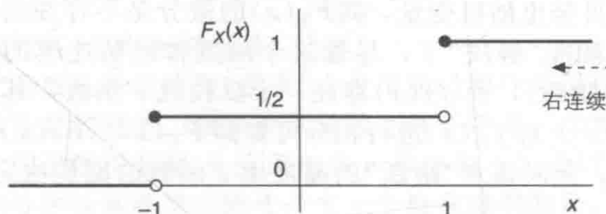


图 3-8 离散贝努里随机变量的分布函数

有两种方法可以表示离散随机变量的输出和概率取值，即利用克罗内克δ函数或者狄拉克δ函数$\delta(x)$。图 3-8 中的分布函数对应的是对称贝努里随机变量，该分布以等概率输出±1。在投掷一枚硬币的时候，就会产生这种类型的随机变量：$H\to X=1$以及$T\to X=-1$。

用克罗内克δ函数来表示离散随机变量的输出会更加方便，此时不存在连续性问题。一个脉冲的高度正好等于输出值x的概率。例 3-10 中，因为$P(X=1)=P(X=-1)=1/2$，所以我们可以写出下面的概率质量函数：

$$p_X[x]=(1/2)(\delta[x-1]+\delta[x+1]) \tag{3-24}$$

(回顾一下，我们用方括号来表示克罗内克δ函数，用圆括号来表示狄拉克δ函数)之所以称作概率质量函数，因为这个函数描述了每个输出值被赋予的概率质量的多少。图 3-9a 给出的是对称贝努里样本的概率质量函数。从δ函数的位置可以看出，输出值(±1)拥有非零概率质量，而且每条线的高度(用实心圆圈表示)给出了每个输出值的概率(1/2)。这种表示方法在概率论基础课程中经常用到，因为它非常直观，而且很清晰地描述了一个离散随机试验的概率值。但是，因为概率质量函数显然不是一个连续函数，并且其支集是Z的一个子集，所以它不能被积分，而积分特性在许多应用中是非常有用的。

对于连续随机变量，将累积分布函数进行微分就可以得到基础概率密度函数，而将概率质量函数进行积分也可以得到累积分布函数(稍后将讨论这个特性)。而对于采用克罗内克δ函数的离散随机变量，该结论不成立。离散问题可以通过狄拉克δ函数$\delta(x)$来解决，该函数在x处的高度为无穷大，宽度为 0，但面积为 1。附录 B 给出了狄拉克δ函数的定义以及其他重要特性。因此，我们可以写出对此贝努里随机变量的概率密度函数：

$$f_X(x)=(1/2)[\delta(x-1)+\delta(x+1)] \tag{3-25}$$

可以看出狄拉克δ函数是通过它的面积来进行定义的：$(1/2)\delta(x)$的面积为 1/2(这并不是狄拉克δ函数的高度)。图 3-9b 给出了对称贝努里随机变量的概率质量函数，在这里我们使用箭头来同克罗内克δ函数进行区分。$F_X(x)$中的"跳跃"是通过单位阶跃函数$u(x)$来进行体现的，其定义如下(如图 3-8 所示)：

$$F_X(x)=(1/2)[u(x+1)+u(x-1)] \tag{3-26}$$

为方便起见，有些图形将在后面给出，在此采用图 3-10，其非连续特性是通过垂线来体现的，而且该函数仍然是右连续的。$F_X(x)$和$f_X(x)$之间的关系将在3.8节进行详细的讨论。

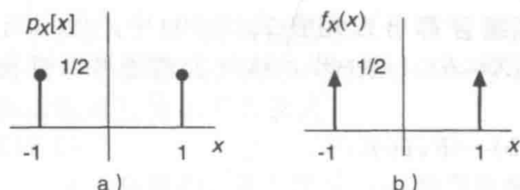


图 3-9 对称贝努里随机变量。a) 基于克罗内尔δ函数的概率质量函数；b) 基于狄拉克δ函数的概率密度函数

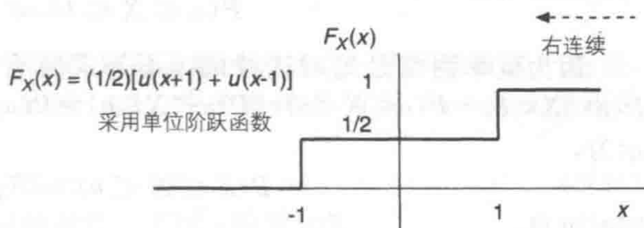


图 3-10 离散随机变量的累积分布函数，该函数的非连续特性是通过垂线来进行体现的(与图 3-8 进行比较)

我们在此强调：对于离散随机变量的问题，其概率密度函数和累积分布函数不是连续的函数。严格来讲，对于贝努里随机变量，其$F_X(x)$的微分是不存在的。但是，通过采用狄拉克δ函数，这个情况就能“解决”了，尽管这个函数曾经是连续的。因为，无论在哪里，$F_X(x)$都有一个断点(阶跃)，而$f_X(x)$存在一个狄拉克δ函数，其面积由断点的高度给出。类似地，在所有的x上对$f_X(x)$进行积分可得到$F_X(x)$。$F_X(x)$一直保持平坦，直到遇到一个狄拉克δ函数，此时该点“跳跃”的高度由δ函数的面积决定(同时也是该点的概率质量)。

例 3-11 设X为一个离散随机变量，输出为$\{1,\cdots,N\}$，定义事件E的概率为$P(E)=C(E)/N$，其中$C(E)=|E|$为计数测度(见第2章)。这个概率空间定义了均匀分布的离散随机变量，该分布在本章后续部分会讲到。假设N是偶数，我们对事件$E=\{$奇数$\cup$数$\leqslant N/2\}$感兴趣，因此有：

$$P(E)=P(\text{奇数})+P(\text{数}\leqslant N/2)-P(\text{奇数}\leqslant N/2) \tag{3-27}$$

根据计数测度可知：

$$C(E)=N/2+N/2-N/4=3N/4\Rightarrow P(E)=3/4 \tag{3-28}$$

利用概率质量函数，也可以写为：

$$P(E)=\sum_{x=1,3,\cdots,N-1}p_X[x]+\sum_{x=1}^{N/2}p_X[x]-\sum_{x=1,3,\cdots,N/2-1}p_X[x] \tag{3-29}$$

由于$p_X[x]=(1/N)I_{\{1,\cdots,N\}}[x]$，它可以从每项求和表达式中提取出来，因此得到式(3-28)的结果。◀

对于等概率输出的有限集合，例3-10中基于计数测度的概率等同于第2章中介绍的$P(E)=|E|/|\Omega|$，我们通过统计输出结果排列或组合的个数来计算E的基数。这种方法在很多问题中非常有用。但是，如果输出结果不是等概率的，或者是更加复杂的随机变量，就有必要通过对概率质量函数求和来计算概率值，同式(3-29)。进一步，还可以用概率质量函数来计算随机变量的各种函数，例如数学期望、方差以及其他矩，在第5章将会对此进行介绍。

3.5 概率密度函数

累积分布函数是对可测空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$上的随机变量$X$在半开半闭区间的一个概率测度，具体见式(3-10)。因为我们对区间感兴趣，应该考虑$F_X(x)$的基础函数，记作概率密度函数$f_X(x)$，对$f_X(x)$在$(-\infty,x]$上进行积分就得到累积分布函数。我们可以通过对基础概率质量函数在一段区间上的积分来计算一个事件的概率，而这段区间是若干个区间的并集。下面给出的这个定理，陈述了概率密度函数的存在性，我们在此不给出证明。

定理 3-2(拉东-尼柯迪姆(Radon-Nikodým)) 如果可测事件空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$上的概率测度$P_X$是绝对连续的，对于随机变量$X$，存在一个非负的函数$f_X(x)$在区间$[a,b]\in\mathcal{R}$上满足：

$$P(a\leqslant X\leqslant b)=\int_a^b f_X(x)\mathrm{d}x \tag{3-30}$$

因为概率测度是绝对连续的，不等式的所有组合都可以用在式(3-30)中，所以有$P(a<X<b)=P(a\leqslant X<b)=P(a<X\leqslant b)=P(a\leqslant X\leqslant b)$。进一步，这些公式还可以被表示为，

$$P(a\leqslant X\leqslant b)=F_X(b)-F_X(a) \tag{3-31}$$

同时可得，

$$F_X(x)=\int_{-\infty}^{x}f_X(x)\mathrm{d}x \tag{3-32}$$

将累积分布函数对x进行求导可得概率质量函数：

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}F_X(x) \tag{3-33}$$

上式也称作 $F_X(x)$的拉东-尼柯迪姆导数。

一般而言，密度函数之所以称作“密度”，是指通过积分可以得到需要的量，在本章中这个量就是某个随机变量的概率。概率密度函数本身并不是一个概率，它是区间[0,1]中一个没有量纲的量。假设随机变量 X 代表电压，单位为 V，此时 X 的概率密度函数的单位为 V^{-1}，当通过积分来计算概率时就可产生一个没有量纲的量。对 X 的概率密度函数在一段区间上进行积分时，我们肯定会得到一个没有量纲的概率值，而且如果这个概率密度函数是正确的，得到的概率值一定会在区间[0,1]中。(同样，在后面我们会考虑功率谱密度 $S_{XX}(f)$，它的单位是 w/Hz。但对 $S_{XX}(f)$在一段频率区间上进行积分时，得到的结果就是功率。)

严格来说，计算拉东-尼柯迪姆导数用于绝对连续概率测度，这意味着 $F_X(x)$是一个绝对连续函数(见定理 3-2)。先前讲到 $F_X(x)$是右连续的。这种差别的存在是因为在实际中 $F_X(x)$可能包含绝对连续、奇异连续和离散分量。一个奇异连续函数的导数几乎处处为0(见例 3-10 中的康托尔函数)。拉东-尼柯迪姆定理是不包括奇异连续函数的，因为它的概率密度函数不能通过式(3-33)计算。前面还提到，在实际情况中不太可能遇到拥有奇异连续部分的累积分布函数，因此我们也不关注这种情况。

另一方面，在许多应用中经常会遇到离散分量，所以离散累积分布函数是值得研究的，同样也包括拥有绝对连续和离散分量的混合累积分布函数。可是，拉东-尼柯迪姆定理也不包括离散的情况，因此式(3-33)的导数定义也是不充分的。3.4 节曾提到，这种情况可以利用狄拉克 δ 函数得到解决，用它来表示 $f_X(x)$的这部分分量，对应的是$F_X(x)$中的断点(阶跃)。这三种类型的累积分布函数及其对应的概率密度函数如图 3-11 所示。对于混合随机变量，连续部分下面的面积加上狄拉克 δ 函数的面积一定等于 1(对于任意随机变量都应该是这样的)。

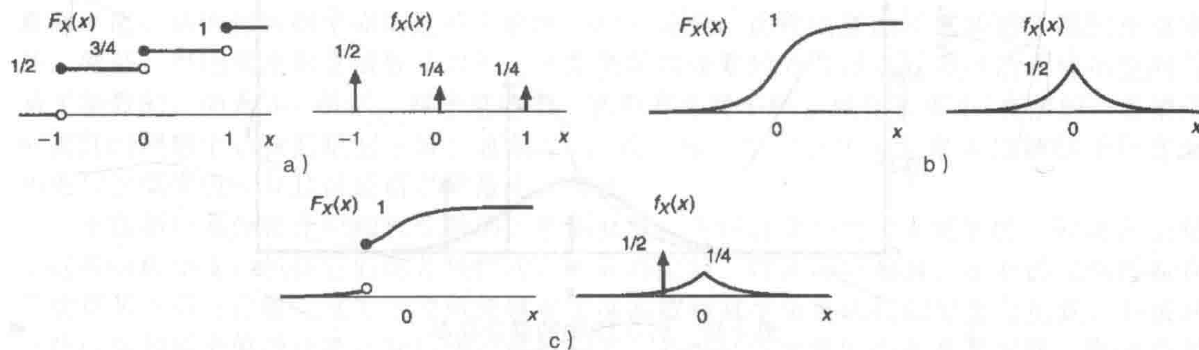


图 3-11　三种类型随机变量的累积分布函数和概率密度函数。a) 离散；b) 连续；c) 混合(包括连续和离散分量)

3.6 混合分布

从式(3-22)的混合概率入手，写出混合累积分布函数的表达式：该函数有若干断点，断点之间是绝对连续的，如图 3-11c 所示。设式(3-22)中的事件 E 为 $X \leqslant x$，那么累积分布函数可写为如下表达式：

$$P(X \leqslant x) = P_{ac}(X \leqslant x) + P_s(X \leqslant x) \tag{3-34}$$

这个分解指出$\mathcal{R}$上的分布函数存在断点的条件。下标 s 的意义为：

$$P_s(X \leqslant x) \triangleq P(X \leqslant x \mid X \in S)P(X \in S) \tag{3-35}$$

其中 S 表示断点处取值的集合(累积分布函数的奇异离散部分)，同样地：

$$P_{ac}(X \leqslant x) \triangleq P(X \leqslant x \mid X \in S^c)P(X \in S^c) \tag{3-36}$$

S^c 是 X 除断点外的支集。(实际上，S 的元素不可能从绝对连续部分被“移除”，因为这些元素在连续部分的概率为 0，它们可以被放进式(3-36)中并且不会影响总的概率。)根据这些定义，式(3-34)可用基于断点的条件概率来表示，可得：

$$F_X(x) = F_{X|S^c}(x|S^c)P(X \in S^c) + F_{X|S}(x|S)P(X \in S) \tag{3-37}$$

其中 $F_{X|S^c}(X|S^c)P(X\in S^c)$是连续部分的累积分布函数，$F_{X|S}(x|S)$是离散部分的累积分布函数，概率 $P(X\in S^c)+P(X\in S)=1$ 提供了一个合适的权重值，使得 $F_X(x)$是一个合法的累积分布函数。对应的概率密度函数为：

$$f_X(x) = f_{X|S^c}(x|S^c)P(X \in S^c) + f_{X|S}(x|S)P(X \in S) \tag{3-38}$$

可以看出，$P(X\in S)$对应的是贝努里随机变量中“成功”的概率，S 中 X 等于 1 表示成功一次。

例 3-12 设 $S=\{0,1,2\}$，$P(X\in S)=0.6$，假设存在下面的等式：

$$f_{X|S^c}(x|S^c) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2) \tag{3-39}$$

$$f_{X|S}(x|S) = (1/8)\delta(x) + (1/4)\delta(x-1) + (5/8)\delta(x-2) \tag{3-40}$$

很显然，第一个概率密度函数是高斯的，第二个概率密度函数是离散的，且由三个输出组成 S。权重 $P(X\in S)=0.6$ 以及 $P(X\in S^c)=0.4$ 确保了 $f_X(x)$是一个合法的概率密度函数，如图 3-12 所示，同时还给出了各自的累积分布函数。

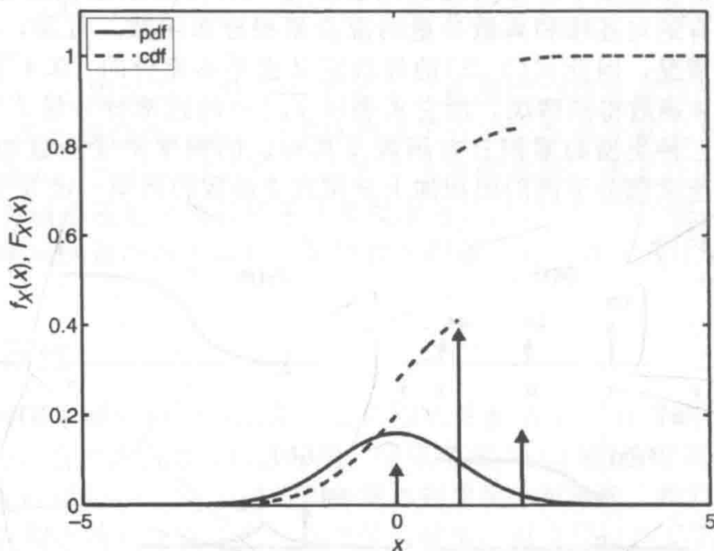


图 3-12 例 3-12 中的混合分布 ◀

给定一个混合累积分布函数，需要判断它的离散部分和连续部分。首先找出函数中的所有断点，这些断点组成了集合 S。每个断点的高度就是它的概率质量(尽管要对其进行比例缩放使得离散概率质量函数之和为 1)。此时，函数剩下的部分就是连续的，通过合适的缩放成为一个合法的概率密度函数。移除断点并不表示将这些输出结果从连续随机变量中排除掉(式(3-36)提到了二者的关系)。如果将连续部分的支集扩展到这些断点，那么这些点就应该被包括在内，只不过概率为 0，而且连续部分中所有的独立点都应该这样处理。在例 3-12 中，混合概率密度函数的支集是$\mathcal{R}$。离散部分的支集是$\{0,1,2\}$，而连续部分的支集仍然是$\mathcal{R}$。

请注意，当两个随机变量相加时得到的式(3-37)并不是一个分布函数。后一种类型的分布函数取决于随机变量的函数，这将在 3.7 节中讨论。混合随机变量实际上是混合分布的特例：

$$F_X(x) = \sum_{n=1}^{N} w_n F_{X_n}(x_n) \tag{3-41}$$

其中$\{F_{X_n}(x_n)\}$是合法的累积分布函数(连续的或离散的)，$\{\omega_n\}$是非负的权重，它使得$\sum_{n=1}^{N}\omega_n = 1$成立。式(3-41)右边的累积分布函数可以是任意形式的，混合分布函数将在第4章学习。

表3-1归纳了本章出现的随机变量的事件空间，同时还包括每个样本空间Ω的基数和相应的σ-域$\mathcal{F}$。Beth数表示的是不同类型无限集合的基数，它们已经在第2章介绍过，表2-4对其进行了总结。通过将连续随机变量的σ-域限制为$\mathcal{B}(\mathcal{R})$，它就与$\mathcal{R}$拥有相同的基数，而幂集(似乎“太大了”)的基数为$\mathcal{P}(\mathcal{R})=\beth_2=P(\mathcal{P}(\mathcal{N}))$。对于离散随机变量(有限的和无限可数的)，样本空间是$\mathcal{R}$的一个子集(N个整数或$\mathcal{Z}$)。事件空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$仍然适用于离散随机变量的累积分布函数$F_X(x)$，且可以支撑实数集合$\mathcal{R}$。严格来说，在这种情况下概率密度函数是不存在的，如果要通过某段区间来描述一个事件的概率，就需要根据黎曼-斯蒂尔切期(Riemann-Stieltjes)积分来计算$F_X(x)$。这里并不打算讨论这种积分，它的数学期望将在第5章讨论，附录D也会对它进行归纳总结。

表3-1 随机变量的事件空间$\{\Omega,\mathcal{F}\}$及其基数

类型	Ω	$\lvert\Omega\rvert$	$\mathcal{F}$	$\lvert\mathcal{F}\rvert$	例子
有限	N个整数	N	$\mathcal{P}(\Omega)$	2^N	贝努里分布
无限可数	$\mathcal{Z}$	$\beth_0=\lvert\mathcal{N}\rvert$	$\mathcal{P}(\Omega)$	$\beth_1=\lvert\mathcal{R}\rvert$	几何分布
连续	$\mathcal{R}$	$\beth_1$	$\mathcal{B}(\Omega)$	$\beth_1$	高斯分布

前面讲过，一个离散随机变量可以包含在一个连续表达式里面，只需要在基础概念密度函数$f_X(x)$中引入狄拉克δ函数即可。事件空间为$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$，通过对$f_X(x)$积分来计算概率值，从而将区间里面的这些$\delta$函数“捕获”进来，这样就考虑了这些感兴趣的离散事件。但是，当用概率质量函数来表示一个离散随机变量的分布时，就可以假设样本空间是基于整数的，如表3-1所示。对于这些Ω，就没有必要限制σ-域比幂集小(在前面一些描述可测性的例题中，我们限制σ-域比幂集小)。进一步，Ω中的所有元素可以被赋予任意分布类型的概率值，并且保证概率质量之和为1。

本章剩余部分将介绍随机变量的一些参数族，先讲连续的然后讲离散的。附录A总结了这些随机变量，包括它们的各种性质，比如各阶矩、特征函数和熵。所有的这些性质在后续章节中都有讲解。这里主要研究概率密度函数和概率质量函数的形态与位置，并讲到一些应用和概率值的计算。利用指示函数来表示各种随机变量的支集非常方便，附录B对其进行了讨论，在此先给出指示函数的定义：

$$I_{(a,b)}(x) \triangleq \begin{cases} 1, & a < x < b \\ 0, & \text{其他} \end{cases} \tag{3-42}$$

同样，还可以在其他类型的区间(闭区间、半开半闭区间)以及独立点上进行定义：

$$I_{(a,b)}[x] \triangleq \begin{cases} 1, & x = a,b \\ 0, & \text{其他} \end{cases} \tag{3-43}$$

它的支集可以用适当的指示函数乘以它的概率密度函数或概率质量函数来表示。

3.7 随机变量的参数模型

在许多应用场合数据是随机的，我们希望推导出一个分布模型，用这个模型可以很准确地表示观测样本。建模的过程有如下几个步骤：

1）给随机变量X的分布选取一个参数族作为模型，该模型能“最好地”代表观测样本。

例如，根据观测样本的直方图，我们可能会选取概率密度函数为$f_X(x)$的高斯分布作为一个合理的模型，因为直方图的形态是“钟形”。

2）根据 X 的观测样本，估计这个分布的参数。对于高斯分布的例子，可以用样本均值 $\overline{X}$ 来估计均值 μ_X，用样本方差 S^2 表示方差 σ_X^2。第 9 章还会讨论许多参数估计方法。

3）当一个参数模型被选定以后，随机变量的各种特性都可以确定，例如特定的概率值或者各阶矩，如偏度和峰度。第 5 章还会讨论矩和其他期望。

4）可以用这个参数模型预测将来的输出，在许多应用场合可以用来做判决，例如后面部分讲到的随机变量的各种类型的族。

例 3-13 图 3-13 给出了 2000 个样本的直方图，x 轴上面每条的宽度为 0.1。这个直方图的外形与指数概率密度函数非常相像，

$$f_X(x) = \lambda\exp(-\lambda x)I_{[0,\infty)}(x) \tag{3-44}$$

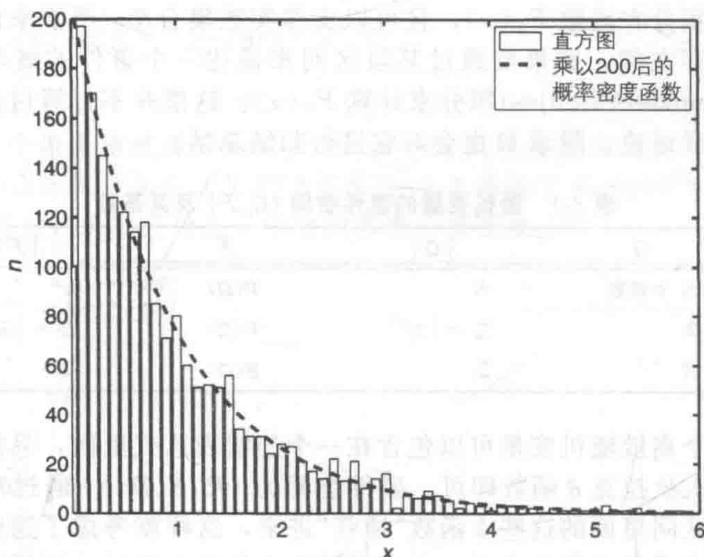


图 3-13 例 3-13 中，2000 个样本的直方图以及概率密度函数（乘以 200）

其中 $\lambda>0$ 是一个参数。指数随机变量的均值为 $\mu_X=1/\lambda$，可以用样本均值来对其估计：

$$\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{3-45}$$

其中$\{x_n\}$表示 N 个实际的输出结果。根据直方图中的 2000 个样本，我们计算出 $\overline{x}=0.9841$，因此估算出 λ 为 1.0162。λ 取这个值的概率密度函数在图 3-13 中也同时给出（不过乘以了 200，这样概率密度函数才会与直方图边缘吻合）。可以看出这个概率密度函数是一个很准确的模型，可以用它来计算特定的概率值，例如：

$$P(X>2) = 1 - F_X(2) = \exp(-2\lambda) \approx 0.1353 \tag{3-46}$$

该概率密度函数在这个例题中是很准确的，因为这些数据都是用 $\lambda=1$ 的指数分布随机数发生器模拟生成的。◀

如果一个分布函数的形态与观测样本的直方图非常吻合，那么这个分布函数就可以作为这个随机变量的有用的模型。下面的这些量值可以用来刻画概率密度函数 $f_X(x)$：

- 均值。概率密度函数的位置：

$$\mu_X \triangleq \int_{-\infty}^{\infty} x\, f_X(x)\mathrm{d}x \tag{3-47}$$

- 方差。概率密度函数的宽度：

$$\sigma_X^2 \triangleq \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\mathrm{d}x \tag{3-48}$$

● 偏度。概率密度函数的不对称特性：

$$\gamma_1 \triangleq \int_{-\infty}^{\infty} [(x-\mu_X)/\sigma_X]^3 f_X(x)\mathrm{d}x = \mu_{c,3}/\sigma_X^3 \tag{3-49}$$

● 峰度。概率密度函数的"拖尾"：

$$\gamma_2 \triangleq \int_{-\infty}^{\infty} [(x-\mu_X)/\sigma_X]^4 f_X(x)\mathrm{d}x - 3 = \mu_{c,4}/\sigma_X^4 - 3 \tag{3-50}$$

其中 $\mu_{c,n}$ 是一个中心矩。各阶矩的定义会在第 5 章中给出，但先假设读者对这两个基本矩比较熟悉：期望和方差。标准差是指方差的正平方根。这里首先介绍它们，然后就可以讨论各种概率密度函数的形状。

有些分布的均值和方差可由不同的参数分别决定，例如高斯随机变量。但是，大部分分布的均值和方差由不止一个参数来共同确定。偏度和峰度是分布参数的一个复杂表达式。γ_1 为正表示分布函数的大部分都在均值以左，γ_1 为负则相反。从 γ_2 的计算公式中减去 3，那么高斯随机变量的峰度为 0。峰度表示与高斯分布相比较时该分布函数拖尾的严重程度。例如，拉普拉斯分布的 $\gamma_2=3$，表示它的拖尾很严重，后面的一些例题将会讲到这个内容。

可以用分位数来描述分布函数的各种概率值。

定义(分位数(quantile))　随机变量 X 的第 m 个分位数 q_m 是指满足下面条件的 x 最小值：

$$F_X(q_m) \geqslant m/M \tag{3-51}$$

其中 $m=1,\cdots,M-1$。

因为 m/M 不是一个整数，所以上面的不等式是用于离散随机变量的。对于连续随机变量，我们取严格的等式 $F_X(q_m)=m/M$。如果 M 是偶数，那么 $q_{M/2}$ 就是中位数 m_e(定义见下面)。$M=4$ 得出的是四分位数，$M=100$ 得出的是百分位数。后面会根据一些特殊的随机变量给出它们的例子。

定义(中位数(median))　对于连续随机变量 X，$f_X(x)$ 的中位数 m_e 是指满足下面等式的取值：

$$\int_{-\infty}^{m_e} f_X(x)\mathrm{d}x = \int_{m_e}^{\infty} f_X(x)\mathrm{d}x \tag{3-52}$$

上面这个表达式与 $F_X(m_e)=1/2$ 是等效的，当 M 是偶数时对应的就是分位数 $q_{M/2}$。对于离散随机变量，这个定义需要进行适当的修改，在第 5 章会讲解。

定义(众数(mode))　一个分布的众数 m_o 是指 $f_X(x)$ 在该处取最大值。

在许多应用中，可以对概率密度函数关于 x 求导来计算众数，导数为 0 时的 x 取值即为 m_o。中位数和众数也会在第 5 章进行讨论。

3.8　连续随机变量

我们在此归纳若干连续随机变量，给出它们各自 $f_X(x)$ 和 $F_X(x)$ 的表达式，并描述它们的一些重要性质。对于有些连续随机变量，它们的累积分布函数没有一个闭式表达式，包括著名的高斯随机变量。对于分布的各种参数，采用下面的符号：

● μ 表示均值(位置参数)。
● σ^2 表示方差(比例参数)。
● σ 表示标准差(比例参数)。
● c 表示一般的位置参数(除均值之外)。
● α 表示一般的比例参数(除标准差之外)。
● λ 表示一般的位置/比例参数。
● r 表示一般的形状参数，或者也表示学生氏 t 分布的自由度。
● α 和 β 一起表示贝塔分布的位置、比例和形态参数。
● m 和 n 表示 F 分布的自由度。
● N 表示卡分布和卡方分布的自由度。

位置参数给出了概率密度函数的“中心位置”(是否真的在中心取决于分布函数族)，尺度参数给出了概率密度函数宽度的一个测量值。这些参数的具体定义将在第 5 章给出。当 r 变化的时候，不存在一个固定的模式来刻画分布函数的形状，概率密度函数的外形可以有很大的变化幅度。本章对每种随机变量都给出了它的概率密度函数图形，对于一组特定类型的参数值，用垂线来指示数学期望 μ_X 的位置(对于离散随机变量一般用虚线)。

在 3.7 节曾提过，本章描述的这些随机变量都很重要，因为它们会出现在各种应用场合。一个具体的概率密度函数往往很有用，因为它的形状是某些“试验”输出结果的非常准确的模型。例如，在许多物理现象中用高斯随机变量来建模经常是非常准确的。工程实际中最重要的例子可能就是噪声，在所有的实际系统中噪声都或多或少地存在。当讨论各种随机变量的时候，概率密度函数是一个很有用的模型，我们可以提供例子告诉大家如何计算概率值以及各种量值，比如数学期望、中位数和众数。

在本章最后，读者可能会意识到所有的概率密度函数和概率质量函数都是 x 的基本函数，许多函数非常简单，而且大部分都有指数部分。如果概率密度函数或者概率质量函数是合法的，那么它一定是单位面积的，即使还包括其他表达式。附录 D 的表 D-1 总结了连续随机变量的概率密度函数，附录 E 的表 E-3 总结了离散随机变量的概率质量函数。由于分布函数的类型太多，还给出了许多不同的函数，目的是为了求累积分布函数，以及其他感兴趣的量，比如矩。为方便读者，附录 B 归纳总结了这些函数以及它们的性质，并给出了不同参数时的图形。

下面将要介绍的这些随机变量没有特定的顺序，但我们尽可能将一些有关联的分布放在一起。为方便查阅，附录 A 中的连续和离散分布是按照字母顺序排列的。

3.8.1　高斯随机变量(正态)

高斯随机变量也称作正态随机变量。它的概率密度函数形状很像一个“钟”：

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/2\sigma^2)I_{(-\infty,\infty)}(x) \tag{3-53}$$

其中均值 $\mu\in\mathcal{R}$，标准差 $\sigma>0$，方差为 σ^2。参数 μ 和 σ 的变化会产生一系列的高斯分布概率密度函数，每个分布的基本形态都相似，差异只在于中心位置和宽度。图 3-14 给出了三个不同的高斯分布概率密度函数曲线。标准正态分布是指 $\mu=0$ 和 $\sigma=1$ 的情况。高斯概率密度函数的支集是所有的 $x\in\mathcal{R}$，用指示函数表示为 $I_{(-\infty,\infty)}(x)$。

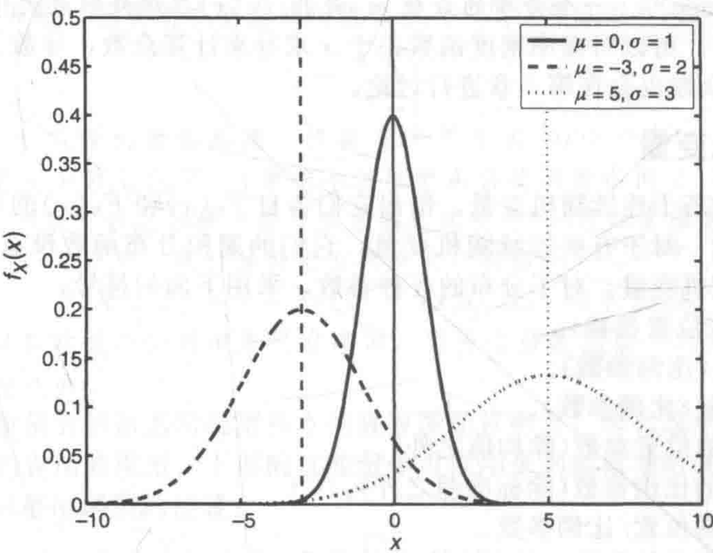


图 3-14　高斯概率密度函数的例子，支集是$\mathcal{R}$，垂线给出了 μ_X 的位置

对于许多物理量和现实应用，高斯随机变量是一个非常有用的模型。尤其是当非常多的数据聚集为一个感兴趣的量的时候，高斯随机变量是非常适用的。第 7 章介绍的中心极

限定理(CLT)指出：大量独立随机变量之和趋向于高斯分布。高斯随机变量模型的应用包括以下内容：

- 电子与机械系统中的噪声。
- 一大群人的物理特征，比如体重和身高。
- 一个大班级中学生的考试分数。

高斯随机变量的累积分布函数为：

$$F_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}\exp\left(-(v-\mu)^2/2\sigma^2\right)\mathrm{d}v \tag{3-54}$$

很遗憾，式(3-54)没有一个闭式表达式。但是，在许多问题中会用到高斯概率密度函数的积分，因此用 Q 函数来简化上面这个表达式。Q 函数基于标准正态分布的，具体定义为：

$$Q(x)\triangleq\frac{1}{\sqrt{2\pi}}\int_{x}^{\infty}\exp\left(-v^2/2\right)\mathrm{d}v \tag{3-55}$$

因此，当 X 是标准高斯随机变量时，根据式(3-55)可知：

$$F_X(x)=1-Q(x) \tag{3-56}$$

有时，用 $\Phi(x)$ 表示标准高斯随机变量的累积分布函数。在本书第 5 章中，已经用 $\Phi_X(\omega)$ 来表示特征函数，因此一般还是采用 Q 函数。附录 B 给出了 Q 函数的查值表，x 从 0 到 3.5。误差函数 $\mathrm{erf}(x)$ 和余误差函数 $\mathrm{erfc}(x)$ 也用来简化高斯概率密度函数的积分表达式，具体请参见附录 B。

为了能在非标准高斯随机变量中使用 Q 函数，需要对积分进行平移变换。(随机变量的一般变换在第 4 章中介绍。)根据下面的变换，任意高斯随机变量 X 的累积分布函数可以通过 Q 函数计算得到，其具体的分布参数为 $\{\mu,\sigma\}$。

$$Y=(X-\mu)/\sigma \tag{3-57}$$

上式中 Y 就是一个标准的高斯随机变量。将这个结果代入式(3-56)可得 X 的累积分布函数为：

$$F_X(x)=1-Q\left(\frac{x-u}{\sigma}\right) \tag{3-58}$$

根据这个表达式，就可以用 $Q(x)$ 表示任意一个高斯随机变量的累积分布函数。

例 3-14 对于一个零均值的高斯随机变量，X 的取值在 3 倍标准差之内的概率为：

$$P(|X|\leqslant 3\sigma)=F_X(3\sigma)-F_X(-3\sigma)=1-2Q(3\sigma/\sigma)\approx 0.9973 \tag{3-59}$$

其中用到了 $F_X(-3)=1-F_X(3)$，而且根据式(3-58)，参数 σ 在运算中被消掉了。◀

例 3-15 证明高斯概率密度函数下方的面积为 1。通过式(3-58)的变换，可以将标准高斯概率密度函数进行如下积分：

$$\int_{-\infty}^{\infty}f_X(x)\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-y^2/2\right)\mathrm{d}y \tag{3-60}$$

这种类型的积分往往可以采用取平方的方法来计算，此时得到一个二重积分，转换到极坐标系下可得：

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-y^2/2\right)\exp(-z^2/2)\mathrm{d}y\mathrm{d}z=\int_{0}^{\infty}\int_{0}^{2\pi}r\exp(-r^2/2)\mathrm{d}r\mathrm{d}\phi \tag{3-61}$$

$y=r\cos(\phi)$，$z=r\sin(\phi)$，所以有 $r^2=y^2+z^2$。该变换的雅克比变换因子为：

$$J=\det\begin{bmatrix}\partial y/\partial r & \partial y/\partial\phi\\ \partial z/\partial r & \partial z/\partial\phi\end{bmatrix}=\det\begin{bmatrix}\cos(\phi) & -r\sin(\phi)\\ \sin(\phi) & r\cos(\phi)\end{bmatrix}=r \tag{3-62}$$

所以可得 $\mathrm{d}y\mathrm{d}z=|J|\mathrm{d}r\mathrm{d}\phi=r\mathrm{d}r\mathrm{d}\phi$。(第 4 章还会介绍雅克比，它可用于随机变量的一般变换，附录 G 也会对其进行总结。)因此，根据部分积分可得：

$$\int_{0}^{\infty}\int_{0}^{2\pi}r\exp(-r^2/2)\mathrm{d}r\mathrm{d}\phi=2\pi \tag{3-63}$$

对式(3-61)取平方根，代入上面的结论可得：

$$\int_{-\infty}^{\infty} \exp\left(-y^2/2\right)\mathrm{d}y = \sqrt{2\pi} \tag{3-64}$$

结论得证。

3.8.2 对数正态随机变量

对高斯随机变量取指数可得对数随机变量。如果 X 服从参数为 $\{\mu,\sigma\}$ 的高斯分布，那么 $Y=\exp(X)$ 服从对数正态分布，概率密度函数为：

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}}\exp(-(\ln(y)-\mu)^2/2\sigma^2)I_{[0,\infty)}(y) \tag{3-65}$$

其中 $\ln(z)\triangleq\log_e(z)$ 表示自然对数。因为指数函数是非负的，所以当 $y\geqslant 0$ 时 $f_Y(y)$ 的取值非零。该随机变量叫做对数正态的原因在于其逆映射：对数正态随机变量 Y 取对数运算后是高斯的。因为 $g(X)=\exp(X)$ 是单调递增的，所以可以根据高斯分布的推导得到对数正态的概率密度函数，这需要用到第 4 章介绍的变换方法。根据逆象 $X=g^{-1}(Y)=\ln(Y)$，可得：

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|I_{[0,\infty)}(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(\ln(y)-\mu)^2/2\sigma^2)\left|\frac{\mathrm{d}\ln(y)}{\mathrm{d}y}\right|I_{[0,\infty)}(y)$$

$$= \frac{1}{y\sqrt{2\pi\sigma^2}}\exp(-(\ln(y)-\mu)^2/2\sigma^2)I_{[0,\infty)}(y) \tag{3-66}$$

图 3-15 为不同取值的 $\{\mu,\sigma\}$ 的对数正态分布的概率密度函数。均值 $\mu_Y=\exp(\mu+\sigma^2/2)$，方差 $\sigma_Y^2=\exp(2\mu+2\sigma^2)-\exp(2\mu+\sigma^2)$。

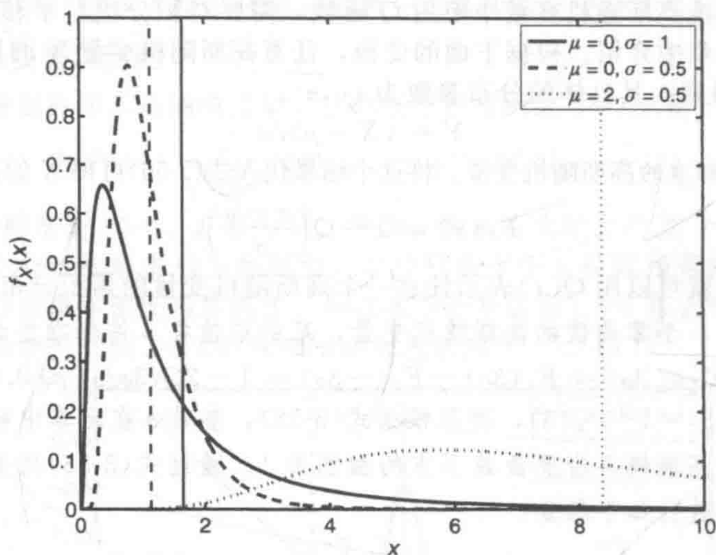


图 3-15 对数正态分布的概率密度函数，支集是 $\mathcal{R}^+$，垂线给出了 μ_X 的位置

对数正态随机变量是一个很有用的模型，尤其是当感兴趣的量取对数后服从高斯分布。根据之前提到的中心极限定理，大量独立随机变量的乘积趋向于对数正态分布。对数正态随机变量模型的应用包括以下内容：

- 地壳中放射性元素的浓度和辐射强度。
- 一句话中单词的个数。
- 一个人接触传染病后发病的时间。

同高斯随机变量类似，对数正态随机变量的累积分布函数没有一个闭式表达式。它可以用误差函数来表示：

$$F_Y(y) = (1/2)\left[1+\operatorname{erf}\left(\frac{\ln(y)-\mu}{\sqrt{2\sigma^2}}\right)\right]I_{[0,\infty)}(y) \tag{3-67}$$

其中误差函数为：

$$\operatorname{erf}(x) \triangleq \frac{2}{\sqrt{\pi}}\int_0^x \exp(-v^2)\,\mathrm{d}v \tag{3-68}$$

附录B给出了余误差函数的查值表，并讨论了它的一些特性。当 $y=\exp(\mu)$ 时，式(3-67)的概率为0.5。

例 3-16 对于一个零均值的对数正态随机变量，Y 的取值在 3σ 之内的概率为：

$$P(Y \leqslant 3\sigma) = (1/2)[1+\operatorname{erf}(\ln(3\sigma)/\sigma\sqrt{2})] \tag{3-69}$$

我们可以看出，与高斯随机变量不同，这个结果与 σ 的取值有关系。请注意，对于正态对数随机变量，σ 并不是它的标准差，它只是一个比例因子。这里我们采用 μ 和 σ(而不是采用一般的位置参数 c 和比例因子 α)只是为了强调对数正态和高斯分布之间的关系。$\ln(\sigma)$ 随着 σ 的增大而增大，因此上面的概率随着 σ 的增大而减小的。例如，$\sigma=1$ 时 $P(Y\leqslant 3\sigma)=0.8640$，而 $\sigma=2$ 时 $P(Y\leqslant 3\sigma)=0.8148$。◀

例 3-17 很容易验证，对数正态随机变量的均值 μ_Y、中位数 m_e 和众数 m_o 的相互关系为：

$$m_o < m_e < \mu_Y \tag{3-70}$$

其中 $m_o=\exp(\mu-\sigma^2)$，$m_e=\exp(\mu)$。◀

3.8.3 反高斯随机变量

反高斯随机变量(Wald)，也称作Wald随机变量，它的概率密度函数为：

$$f_X(x) = \sqrt{\frac{\alpha}{2\pi x^3}}\exp(-\alpha(x-\mu)^2/2\mu^2 x)I_{[0,\infty)}(x) \tag{3-71}$$

其中 $\alpha>0$ 为比例因子，$\mu\in\mathcal{R}$ 为均值。方差 $\sigma_X^2=\mu^3/\alpha$。由于累积分布函数没有闭式表达式，与高斯随机变量类似，它也可以写成 Q 函数的形式：

$$F_X(x) = 1-Q((x/\mu-1)\sqrt{\alpha/x})+\exp(2\alpha/\mu)Q((x/\mu+1)\sqrt{\alpha/x}) \tag{3-72}$$

还可以写成误差函数的形式：

$$F_X(x) = (1/2)[1+\operatorname{erf}((x/\mu-1)\sqrt{\alpha/2x}]+(1/2)\exp(2\alpha/\mu)[1-\operatorname{erf}(x/\mu+1)\sqrt{\alpha/2x})] \tag{3-73}$$

概率密度函数的一些例子见图3-16。称作反高斯随机变量的原因在于其累积量母函数(见第5章)和高斯随机变量之间的反比关系。该变量的一些应用如下：

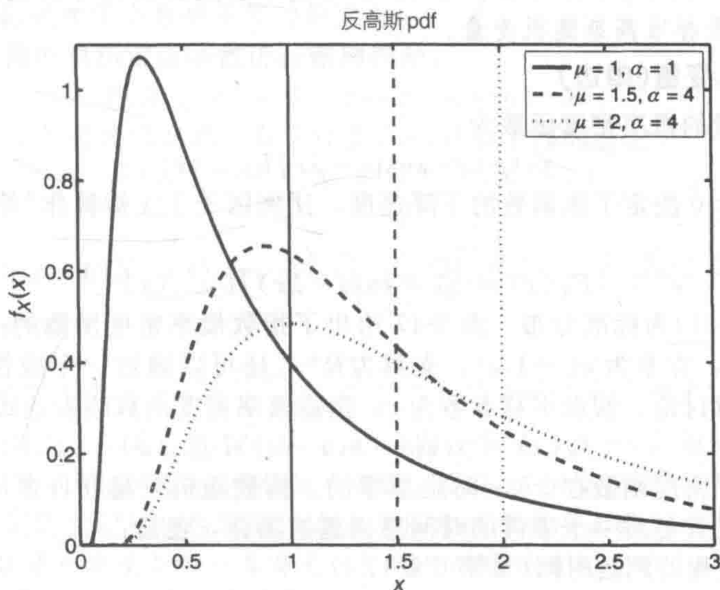


图3-16 反高斯pdf的示例。支集是 R^+，垂线给出了 μ_X 的位置

- 描述随机扰动的时间，每个等级上的偏移都是正的(称作首次通过时间)。

- 给风速建模，尤其是很少出现低风速的场合。
- 各行业工人罢工的持续时间。

例 3-18 对 Wald 随机变量 X 取倒数得到随机变量 $Y=1/X$，其概率密度函数为：

$$f_Y(y)=\sqrt{\frac{\alpha}{2\pi y}}\exp(-\alpha y/2-\alpha/2\mu^2 y+\alpha/\mu)I_{[0,\infty)}(y) \tag{3-74}$$

上式通过累积分布函数可以很容易得到：

$$F_Y(y)=P(1/X\leqslant y)=P(X\geqslant 1/y)=1-F_X(1/y) \tag{3-75}$$

而两边同时乘以 $X>0$ 并不会影响不等式的方向，因此可得：

$$f_Y(y)=f_X(1/y)(1/y^2) \tag{3-76}$$

将 $x=1/y$ 代入式(3-71)，将结果乘以 $1/y^2$，整理指数函数表达式，得到式(3-74)。◀

例 3-19 对于一对一映射 $Y=\sqrt{\alpha}(X-\mu)/\sqrt{X}\mu$，$Y$ 的概率密度函数为如下表达式(见习题 4-26)，支集为 $y\in\mathcal{R}$：

$$f_Y(y)=\frac{1}{\sqrt{2\pi}}(1-y/\sqrt{4\alpha/\mu+y^2})\exp(-y^2/2) \tag{3-77}$$

这个结果非常有趣，它是标准高斯随机变量概率密度函数与一个表达式的乘积，该表达式是常数 1 和奇函数 $y/\sqrt{4\alpha/\mu+y^2}$ 之差。根据这个结果，很容易验证 $f_Y(y)$ 下方的面积为 1，因为奇函数部分的积分为 0。考虑下面这个变换 $Z=|Y|$，其累积分布函数为：

$$\begin{aligned}P(Z\leqslant z)=P(-z\leqslant Y\leqslant z)&=\frac{1}{\sqrt{2\pi}}\int_{-z}^{z}(1-y/\sqrt{4\alpha/\mu+y^2})\exp(-y^2/2)\mathrm{d}y\\&=\frac{1}{\sqrt{2\pi}}\int_{-z}^{z}\exp(-y^2/2)\mathrm{d}y=\frac{2}{\sqrt{2\pi}}\int_{0}^{z}\exp(-y^2/2)\mathrm{d}y\end{aligned} \tag{3-78}$$

其中奇函数部分的积分为 0。对 z 求导，得到概率密度函数表达式为：

$$f_Z(z)=\frac{2}{\sqrt{2\pi}}\exp(-z^2/2) \tag{3-79}$$

上式对应的是零均值单位方差的标准半-正态分布(稍后会介绍)。根据 $Z=|Y|$ 可知 Z 的支集是 $\mathcal{R}^+$，同样，$f_Z(z)$ 下方的面积为 1。Z 的概率密度函数也可以通过变换 $Z=|U|$ 类似计算，其中 U 是标准高斯随机变量。◀

3.8.4 指数随机变量(单边)

指数随机变量的概率密度函数为：

$$f_X(x)=\lambda\exp(-\lambda x)I_{[0,\infty)}(x) \tag{3-80}$$

位置/比例参数 $\lambda>0$ 决定了该函数的下降速度，比例因子 $1/\lambda$ 被称作“等待时间”。累积分布函数为：

$$F_X(x)=[1-\exp(-\lambda x)]I_{[0,\infty)}(x) \tag{3-81}$$

当 $\lambda=1$ 时，式(3-81)为标准分布。图 3-17 给出了指数概率密度函数的一些例子。该分布的均值为 $\mu_X=1/\lambda$，方差为 $\sigma_X^2=1/\lambda^2$。支集为 $\mathcal{R}^+$，还可以通过一个线性变换对其进行平移，这将在第 4 章讨论。假设平移参数为 c，指数概率密度函数的表达式为：

$$f_X(x)=\lambda\exp(-\lambda(x-c))I_{[c,\infty)}(x) \tag{3-82}$$

指示函数确保概率密度函数在 $x\geqslant c$ 时是非零的。指数随机变量在许多应用场合是一个非常有用的模型，尤其是对一个事件的时间感兴趣的场合，包括：

- 泊松随机过程的到达时间(见第 6 章)。
- 放射性粒子的衰变。
- 一个电子元器件的失效时间。

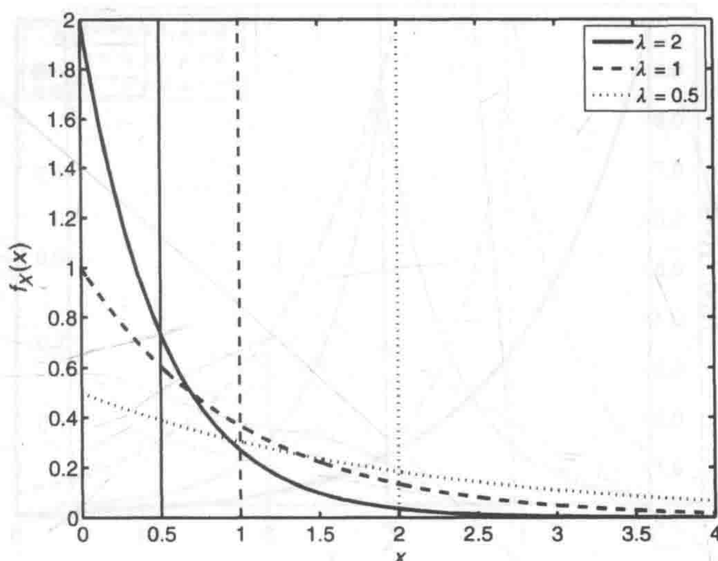


图 3-17 指数概率密度函数的例子，支集是$\mathcal{R}^+$，垂线给出了 μ_X 的位置

例 3-20 指数随机变量的中位数可以通过计算 $F_X(0.5)$得到：

$$2(1-\exp(-\lambda m_e)) = 1 \Rightarrow m_e = \ln(2)/\lambda \tag{3-83}$$

式(3-80)的众数始终为 0，因此有 $m_o < m_e < \mu_X$。◀

例 3-21 指数随机变量是伽马随机变量的一个特例，此时伽马随机变量的比例因子为 $\alpha=1/\lambda$，形状参数 $r=1$。伽马随机变量稍后会进行介绍。通过适当的变换，指数随机变量可以与其他类型的随机变量有关联，包括极值分布(在后面也将介绍)：

$$Y = c - \alpha\ln(X) \tag{3-84}$$

其中 $c\in\mathcal{R}$表示位置参数，$\alpha>0$ 为比例因子。对累积分布函数进行整理：

$$\begin{aligned} P(Y \leqslant y) &= P(c-\alpha\ln(X) \leqslant y) \\ &= P(X \geqslant \exp((c-y)/\alpha)) = \exp(-\lambda\exp((y-c)/\alpha)) \end{aligned} \tag{3-85}$$

式(3-85)最后一行就是 $\lambda=1$ 时极值分布的累积分布函数。请注意，在式(3-85)中由于 $\ln(\cdot)$是单调的，因此并不会影响不等式的方向。◀

对指数随机变量的累积密度函数进行整理得到：

$$P(X > x) = 1 - F_X(x) = \exp(-\lambda x) \tag{3-86}$$

从式(3-86)可以看出 X 是无记忆的。对于任意 $b\geqslant 0$，有下面的结论：

$$P(X > x+b \mid X > b) = P(X > x) \tag{3-87}$$

根据条件概率有：

$$P(X > x+b \mid X > b) = \frac{P(X > x+b, X > b)}{P(X > b)} = \frac{P(X > x+b)}{P(X > b)} \tag{3-88}$$

代入式(3-86)可得：

$$P(X > x+b \mid X > b) = \frac{\exp(-\lambda(x+b))}{\exp(-\lambda b)} = \exp(-\lambda x) = P(X > x) \tag{3-89}$$

这正是我们期待的结论。如果已知 $X>b$，$b\geqslant 0$，那么 X 大于 $x+b$ 的概率与$P(X>x)$是相同的，其中 X 是原始的(未平移的)指数随机变量。图 3-18 形象地表示了这种无记忆特性。

接下来，我们证明无记忆特性实际上定义了指数概率密度函数。

定理 3-3 指数随机变量是唯一具有无记忆特性的连续随机变量。

证明： 对于如下的条件累积分布函数：

$$P(X \leqslant x+b \mid X > b) = \frac{P(b < X \leqslant x+b)}{P(X > b)} \tag{3-90}$$

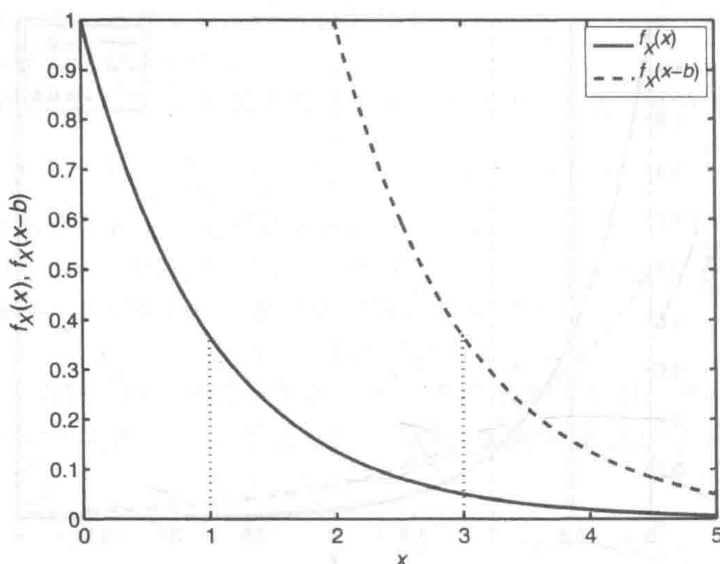


图 3-18　指数随机变量的无记忆特性，$\lambda=1$。两条曲线到各自垂线下方的面积是相同的。以 $X>b>0$ 为前提条件等同于将 $f_X(x)$ 右移 b，本例中 $b=2$

根据无记忆特性，

$$\frac{F_X(x+b)-F_X(b)}{1-F_X(b)}=F_X(x)-F_X(0) \tag{3-91}$$

其中 $F_X(x)$ 就是我们需要计算的累积分布函数，$F_X(0)=0$ 是边界条件。根据积分运算，我们将等式两边同时除以 $(x+b)-b$，并令 $x\to 0$ 得到 $F_X(x)$ 的在 $x=b$ 和 $x=0$ 处的导数：

$$\frac{1}{1-F_X(b)}F'_X(b)=F'_X(0)\Rightarrow\frac{\mathrm{d}}{\mathrm{d}x}F_X(x)=[1-F_X(x)]F'_X(0) \tag{3-92}$$

为方便，我们已将上式中的变量 b 替换为 x。上式是一个一阶微分方程表达式，最终的结果一定是指数形式：$1-F_X(x)=a\exp(cx)$，$a\neq 0$。将这个表达式代入式(3-92)，求导后可得：

$$-ac\exp(cx)=a\exp(cx)F'_X(0)\Rightarrow c=-F'_X(0) \tag{3-93}$$

对于连续随机变量，其累积分布函数的导数是正的，因此可知 $c<0$。代入 $c=-\lambda$ 到表达式 $1-F_X(x)$ 中可得：

$$1-F_X(x)=a\exp(-\lambda x)\Rightarrow F_X(x)=1-a\exp(-\lambda x) \tag{3-94}$$

对上式进行求导可得 $a=\lambda$ 时的概率密度函数(即式(3-80))，命题得证。

3.8.5　拉普拉斯随机变量(双边指数)

拉普拉斯随机变量，也称作双边指数随机变量，概率密度函数为：

$$f_X(x)=(1/2\alpha)\exp(-|x-\mu|/\alpha)I_{(-\infty,\infty)}(x) \tag{3-95}$$

$\mu\in\mathcal{R}$ 表示均值，比例因子 $\alpha>0$ 决定了该函数的下降速度。拉普拉斯概率密度函数是通过对式(3-82)的指数概率密度函数进行平移变换得到的：添加它的“镜像”$c=\mu$，设 $\alpha=1/\lambda$，对结果乘以 1/2 的比例因子，保证双边概率密度函数下方的面积为 1。不同于指数概率密度函数中的 λ，α 是一个严格的比例因子并且与 μ 是独立的。对拉普拉斯随机变量 X 取绝对值 $|X|$ 运算，并且代入 $\alpha=1/\lambda$ 和 $\mu=0$，X 就退化为标准指数随机变量(不要将 $|X|$ 与集合的基数相混淆)。拉普拉斯概率密度函数的例子如图 3-19 所示，方差为 $\sigma_X^2=2\alpha^2$。

拉普拉斯随机变量在许多场合是一个非常有用的模型，尤其是当概率密度函数存在“大拖尾现象”时。从图 3-20 可以看出，拉普拉斯和高斯概率密度函数的外形非常相似，不过拉普拉斯分布的概率密度函数在 $x=\mu$ 处有一个尖峰，而且它下降得更缓慢一些，这是拉普拉斯分布中的 $\exp(-x)$ 与高斯分布中 $\exp(-x^2)$ 的缘故。拉普拉斯随机变量的应用包括：

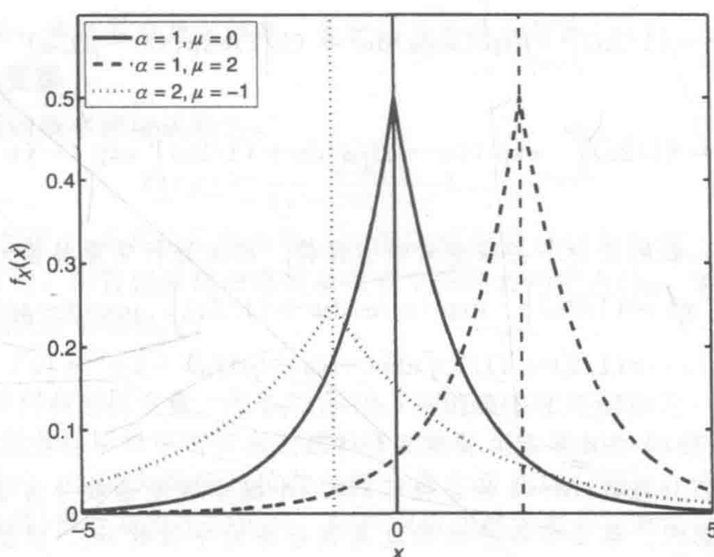


图 3-19　拉普拉斯概率密度函数的例子，支集是$\mathcal{R}$，垂线给出了 μ_X 的位置

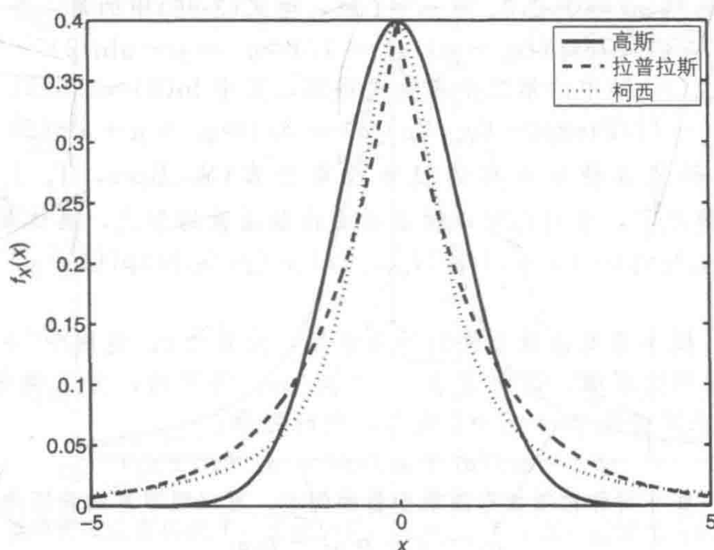


图 3-20　高斯($\sigma=1$)，拉普拉斯($\alpha=\sqrt{\pi/2}$)，柯西($\alpha=\sqrt{2/\pi}$)的概率密度函数，位置参数都相同($\mu=c=0$)。选取适当的比例因子，使得三个概率密度函数的最大值相同

- 脉冲噪声中的信号检测，需要比高斯噪声更严重的拖尾。
- 编码和解码中语音信号的建模。
- 将非对称拉普拉斯分布(见例 3-24)用于财经数据建模，比如利率。

拉普拉斯随机变量的累积分布函数为：

$$F_X(x)=\begin{cases}(1/2)\exp((x-\mu)/\alpha), & x<\mu\\ 1-(1/2)\exp(-(x-\mu)/\alpha), & x\geqslant\mu\end{cases}\tag{3-96}$$

$$=(1/2)(1+\operatorname{sgn}(x-\mu)[1-\exp(-|x-\mu|/\alpha)])I_{(-\infty,\infty)}(x)\tag{3-97}$$

其中符号函数为(见附录 B)：

$$\operatorname{sgn}(x)\triangleq\begin{cases}1, & x\geqslant 0\\ -1, & x<0\end{cases}\tag{3-98}$$

例 3-22　拉普拉斯分布的累积分布函数推导过程如下。对 $x\leqslant\mu$ 的部分直接积分，通过变换替换 $w\triangleq v-\mu$：

$$P(X\leqslant x)=(1/2\alpha)\int_{-\infty}^{x}\exp((v-\mu)/\alpha)\mathrm{d}v$$

$$=(1/2\alpha)\int_{-\infty}^{x-\mu}\exp\,(w/\alpha)\mathrm{d}w=(1/2)\exp((x-\mu)/\alpha) \tag{3-99}$$

对于 $x>\mu$ 的部分：

$$P(X\leqslant x)=(1/2\alpha)\int_{-\infty}^{\mu}\exp\,((v-\mu)/\alpha)\mathrm{d}v+(1/2\alpha)\int_{\mu}^{x}\exp\,(-(v-\mu)/\alpha)\mathrm{d}v \tag{3-100}$$

将积分拆成两部分，目的是为了去除绝对值运算符。再次进行变量替换 $w\triangleq v-\mu$，可得：

$$\begin{aligned}P(X\leqslant x)&=(1/2\alpha)\int_{-\infty}^{0}\exp\,(w/\alpha)\mathrm{d}w+(1/2\alpha)\int_{0}^{x-\mu}\exp\,(-w/\alpha)\mathrm{d}w\\&=(1/2)-(1/2)[\exp(-(x-\mu)/\alpha)-1]\\&=1-(1/2)\exp(-(x-\mu)/\alpha)\end{aligned} \tag{3-101}$$

将这个结果与式(3-99)合并在一起，得到式(3-97)。◀

例 3-23 三个四分位数($M=4$ 的分位数)将一个概率密度函数分为了四个部分，每部分都是四分之一的面积。第 2 个四分位数就是该分布的中位数 m_e。很显然，由于拉普拉斯分布是对称的，因此 $m_e=\mu$。四分位数是通过求解连续随机变量的累积分布函数得到的，即 $F_X(q_m)=m/4, m=1,2,3$。当 $m=1$ 时，由式(3-96)中的第一个表达式得到：

$$(1/2)\exp((q_1-\mu)/\alpha)=1/4\Rightarrow q_1=\mu-\alpha\ln(2) \tag{3-102}$$

当 $m=3$ 时，根据式(3-96)中的第二个表达式得到，其中 $\ln(2)\approx 0.6931$。

$$1-(1/2)\exp(-(q_3-\mu)/\alpha)=3/4\Rightarrow q_3=\mu+\alpha\ln(2) \tag{3-103}$$ ◀

例 3-24 在此研究各种非对称的拉普拉斯分布(S. Kotz, T. J. Kozubowski, K. Podgorski)。一般情况下，它们总可以被表示成指数函数的形式，具体如下：

$$f_X(x)=(p_1/\alpha_1)\exp(-(x-c)/\alpha_1)I_{[c,\infty)}(x)+(p_2/\alpha_2)\exp((x-c)/\alpha_2)I_{(-\infty,c)}(x) \tag{3-104}$$

正数 $\{\alpha_1,\alpha_2\}$ 给出了概率密度函数下降的不同速度，众数为 c。这两个“半边概率密度函数”在 $x=c$ 处必须有相同的取值，因此有 $p_1/\alpha_1=p_2/\alpha_2$。同样的，为了确保整体是一个有效的概率密度函数，必须保证 $p_1+p_2=1$ 成立，此时可得：

$$p_1=\alpha_1/(\alpha_1+\alpha_2),\ p_2=\alpha_2/(\alpha_1+\alpha_2) \tag{3-105}$$

图 3-21 给出了一些非对称拉普拉斯随机变量的例子。可以很容易的验证均值为(见习题 5-29)：

$$\mu_X=c+p_1\alpha_1-p_2\alpha_2 \tag{3-106}$$

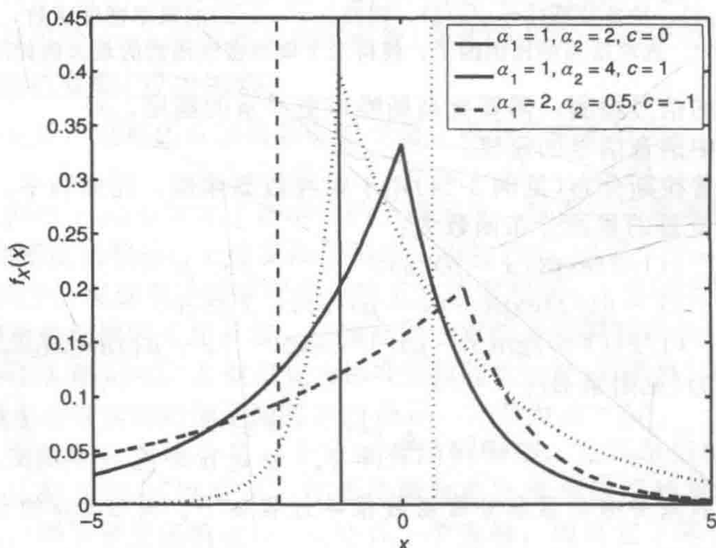


图 3-21 非对称概率密度函数(式(3-104))的一些例子，支集是$\mathcal{R}$，垂线给出了 μ_X 的位置

当且仅当 $\alpha_1=\alpha_2$ 时，均值与众数 c 相等，此时的分布是对称的。◀

3.8.6 柯西随机变量

柯西随机变量的概率密度函数为：

$$f_X(x)=\frac{\alpha/\pi}{(x-c)^2+\alpha^2}I_{(-\infty,\infty)}(x) \tag{3-107}$$

其中 $\alpha>0$ 为比例因子，$c\in\mathcal{R}$是位置参数。柯西随机变量也称作柯西－洛伦兹随机变量。在第 5 章将会学习到，尽管柯西概率密度函数有个很明显的中点(用 c 来表示)，但均值 μ_X 不存在。相应的累积分布函数为：

$$F_X(x)=[1/2+(1/\pi)\arctan((x-c)/\alpha)]I_{(-\infty,\infty)}(x) \tag{3-108}$$

$c=0$ 和 $\alpha=1$ 为标准柯西随机变量。图 3-22 给出了柯西概率密度函数的一些例子，在图 3-20 中还将其与高斯和拉普拉斯概率密度函数进行了对比。可以看出，柯西概率密度函数是三种分布中下降最慢的，它的大拖尾与拉普拉斯分布非常相似。因此，在“离群数据”(x 中非常大的正值和负值)很容易发生的场合，柯西随机变量的模型比高斯随机变量更准确。

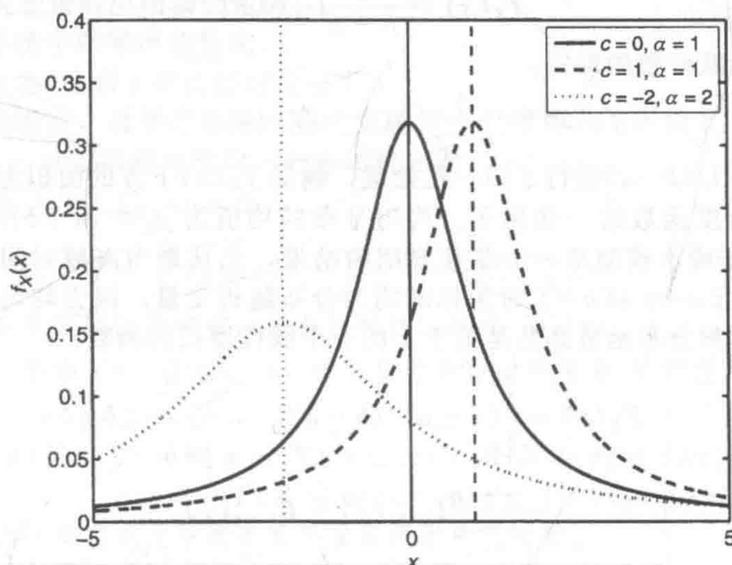


图 3-22 柯西概率密度函数的例子，支集为$\mathcal{R}$，因为 μ_X 不存在，垂线给出的是 m_e 的位置

例 3-25 一个函数的拐点发生在其二阶导数改变符号的时候。对于柯西分布有：

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}f_X(x)=\frac{-2\alpha/\pi}{[(x-c)^2+\alpha^2]^2}+\frac{-6(\alpha/\pi)x}{(x-c)^2+\alpha^2} \tag{3-109}$$

令上式为 0，对方程关于 x 求解，得到两个拐点：

$$x=c\pm\alpha/\sqrt{3} \tag{3-110}$$

将结果代入累积分布函数可得：

$$F_X(c\pm\alpha/\sqrt{3})=1/2+(1/\pi)\arctan(\pm 1/\sqrt{3})=\{1/3,2/3\} \tag{3-111}$$

高斯随机变量的对应的值就已经在这个区间之外了(见习题 3-20)，这很好地说明了柯西分布有更严重的拖尾。◀

柯西随机变量的应用场合包括：

- 从一个点源发射出的粒子击中一个固定直线的分布。
- 如果 X 和 Y 是独立的标准高斯随机变量，那么 $Z=X/Y$ 是标准柯西随机变量(见第 4 章)。
- 地震几何学的表示。

尽管柯西分布的均值以及其他阶矩都不存在，但柯西分布的中位数和众数是存在的，二者都等于 c。

例 3-26 为不失一般性，假设 $c=0$，所以概率密度函数的中心在 $x=0$ 处。柯西分布的均值可以计算如下：

$$\int_{-\infty}^{\infty} x \frac{\alpha/\pi}{x^2+\alpha^2} \mathrm{d}x = (\alpha/2\pi)\log(x^2+\alpha^2)\Big|_{-\infty}^{\infty} = \infty - \infty \tag{3-112}$$

可以看出，得到的结果是没有定义的(并不是0)。(请参看附录E，包括0和∞的各种等式。) ◀

例 3-27 柯西随机变量的分位数计算如下，$m=1,\cdots,M-1$：

$$1/2+(1/\pi)\arctan((q_m-c)/\alpha) = m/M \Rightarrow q_m = c+\alpha\tan((2m-M)\pi/2M) \tag{3-113}$$

当 $c=0$，$\alpha=1$ 时的十分位数($M=10$)为 −3.0777，−1.3764，−0.7265，−0.3249，0，0.3249，0.7265，1.3764 和 3.0777。 ◀

3.8.7 连续均匀随机变量

均匀分布随机变量的概率密度函数在$\mathcal{R}$的一个有限支集合上是平坦的(矩形的)。在区间 $[a,b]$ 上，$a,b\in\mathcal{R}$ 且 $a<b$，均匀分布的概率密度函数为：

$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x) \tag{3-114}$$

也可以写成单位阶跃函数的形式：

$$f_X(x) = \frac{1}{b-a}[u(x-a)-u(x-b)] \tag{3-115}$$

式(3-15)中用 $1/(b-a)$ 进行了归一化处理，确保 $f_X(x)$ 下方的面积为 1。图 3-23 给出了均匀分布概率密度函数的一些例子。均匀分布的均值为 $\mu_X=(a+b)/2$，方差为 $\sigma_X^2=(b-a)^2/2$，这对于噪声模型是一个非常有用的结果，尤其是当离散时间信号被量化的时候(见第 10 章)。当 $a=0$ 和 $b=1$ 时为标准均匀分布随机变量。因为均匀分布概率密度函数是平坦的，其累积分布函数必然是关于 x 的一个线性增长的函数：

$$F_X(x) = \begin{cases} 0, & x<a \\ (x-a)/(b-a), & a\leqslant x<b \\ 1, & x\geqslant b \end{cases}$$
$$= \frac{x-a}{b-a} I_{[a,b]}(x) + I_{(b,\infty)}(x) \tag{3-116}$$

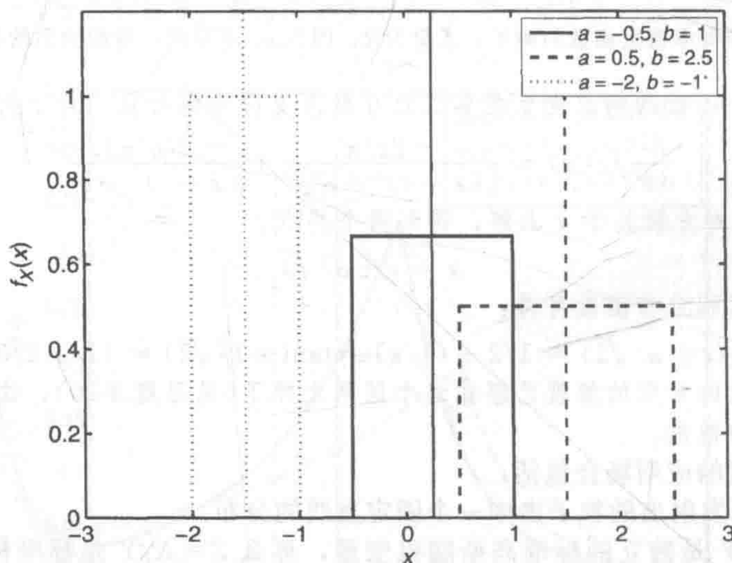


图 3-23　连续均匀分布的概率密度函数，支集是$[a,b]$，垂线指出了 μ_X 的位置

在表达式中一定要加入指示函数 $I_{(b,\infty)}(x)$，这样才能保证 x 超出 $f_X(x)$ 的支集时 $F_X(x)=1$。累积分布函数也可以表示成斜坡函数的形式：

$$F_X(x)=\frac{1}{b-a}[R(x-a)-R(x-b)]I_{[a,\infty)}(x) \tag{3-117}$$

其中斜坡函数的定义为(见附录 B)：

$$R(x)\triangleq\begin{cases}x, & x\geqslant 0\\ 0, & x<0.\end{cases} \tag{3-118}$$

概率密度函数是平坦的，意味着$[a,b]$中所有宽度相同的子区间具有相同的概率：对于$x_1<x_2$，$x_3<x_4$，并且x_1，x_2，x_3，$x_4\in[a,b]$，如果$L([x_1,x_2])=L([x_3,x_4])$，$L(\cdot)$表示的是$L$测度，那么有：

$$P(x_1\leqslant X\leqslant x_2)=P(x_3\leqslant X\leqslant x_4) \tag{3-119}$$

均匀分布子区间的概率并不取决于该子区间在随机变量支集里的具体位置。当所有的输出都是“等随机的”应用场合，均匀分布随机变量是一个非常有用的模型。(这个表达式对于离散均匀分布随机变量具有更多的含义，所有输出结果都有相同的概率，即有限数量的输出结果是“等可能的”。)

均匀分布随机变量的应用场合包括：

- 一个数字系统中对噪声的量化。
- 随机数发生器(在第 1 章已经讨论过了)。
- 在第 1 章中讲过，对于任意连续随机变量X及其累积分布函数$F_X(x)$，概率积分变换$Y=F_X(x)$可以得到标准均匀分布随机变量。

例 3-28 对于$c,d\in[a,b]$可得：

$$\int_c^d f_X(x)\mathrm{d}x=\frac{d-c}{b-a}=\frac{L([c,d])}{L([a,b])} \tag{3-120}$$

这个结果证明了等宽度区间具有相同概率的特性，即式(3-119) ◀

例 3-29 通过变换$Y=-\ln(X)/\lambda$，标准均匀分布随机变量X和指数随机变量具有如下关系，其中$a=0$，$b=1$：

$$\begin{aligned}P(Y\leqslant y)&=P(-\ln(X)/\lambda\leqslant y)=P(X>\exp(-\lambda y))\\&=1-P(X\leqslant\exp(-\lambda y))=1-\exp(-\lambda y)\end{aligned} \tag{3-121}$$

上面最后一个表达式其实就是指数随机变量的概率密度函数。 ◀

3.8.8 三角随机变量

三角随机变量的概率密度函数为：

$$f_X(x)=\frac{2}{b-a}\left[\frac{x-a}{c-a}I_{[a,c]}(x)+\frac{b-x}{b-c}I_{(c,b]}(x)\right][u(x-a)-u(x-b)] \tag{3-122}$$

其中$a,b,c\in\mathcal{R}$且$a<c<b$，支集为$[a,b]$。通过观测可知$f_X(x)$是一个通用的三角函数，上升部分的斜率$(1/(c-a))$不一定与下降部分的斜率$(1/(b-c))$相等。均值为$\mu_X=(a+b+c)/3$，方差为$\sigma_X^2=(a^2+b^2+c^2-ab-ac-bc)/18$。图 3-24 给出了三角分布概率密度函数的一些例子。

累积分布函数为：

$$F_X(x)=\frac{1}{b-a}\left[\frac{(x-a)^2}{c-a}I_{[a,c]}(x)+\left(b-a-\frac{(b-x)^2}{b-c}\right)I_{(c,b]}(x)\right]+I_{(b,\infty)}(x) \tag{3-123}$$

在式(3-123)中引入$I_{(b,\infty)}(x)$，这样确保当x超出$f_X(x)$的支集时$F_X(x)=1$。$F_X(x)$在$x=c$处有一个拐点，在这里$F_X(x)$从凸函数变为了凹函数(见附录 A 中累积分布函数的例子)。$F_X(x)$在拐点处的取值为$(c-a)/(b-a)$，当且仅当$c=(a+b)/2$时$F_X(x)$取 1/2，此时的三角函数关于c对称。

三角分布随机变量的应用包括：

- 当一个分布的最大值、最小值以及“最有可能的”取值都是已知的，对该分布进行建模(分别对应于a、b、c的估计)。

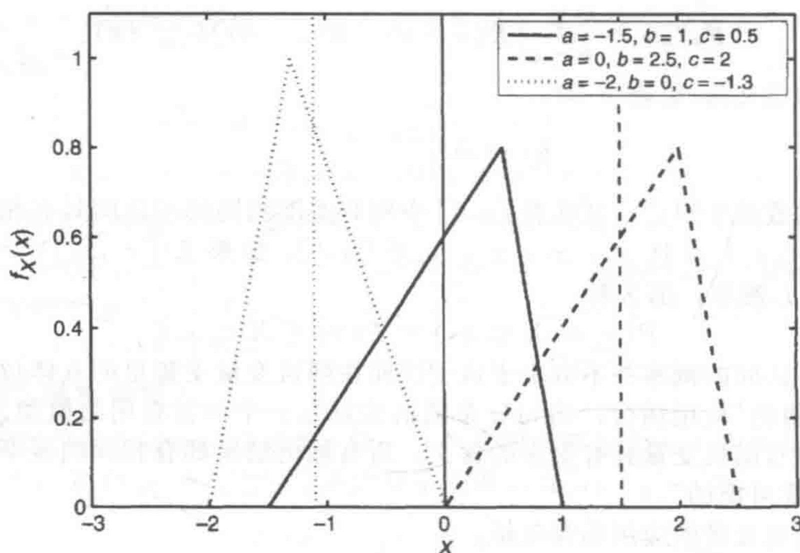


图 3-24 三角分布概率密度函数的一些例子，支集为$[a,b]$，垂线标出了 μ_X 的位置

- 对 beta 分布进行近似(后面章节将会具体讲到)。
- 数字信号的抖动显示，目的是为了降低在量化过程中引起的失真，并且与原始信号是相互关联的。

例 3-30 三角随机变量的斜率系数可以通过它的三阶中心矩 $\mu_{c,3}$(定义见第 5 章)和方差 σ_X^2(见附录 A)推导得到：

$$\gamma_1 = \mu_{c,3}/\sigma_X^3 = (\sqrt{2}/5)\,\frac{(a+b-2c)(2a-b-c)(a-2b+c)}{(a^2+b^2+c^2-ab-ac-bc)^{3/2}}. \tag{3-124}$$

从分子部分可以看出，当且仅当 $c=(a+b)/2$ 时 $\gamma_1=0$。因为 $a<c<b$，分子中的另外两项不可能为 0。当 $a=0$ 和 $b=1$ 时：

$$\gamma_1 = (\sqrt{2}/5)\,\frac{(1-2c)(c+1)(2-c)}{(c^2-c+1)^{3/2}} \tag{3-125}$$

结果如图 3-25 所示。当 $c>1/2$ 时，三角分布的斜率为负，意思是三角形此时往右边倾斜("依靠")；类似的，当 $c<1/2$ 时 $\gamma_1>0$，此时的三角形往左边倾斜。图 3-25 同时还给出了 $c=1/10$ 和 $c=3/4$ 时的概率密度函数，显示了斜率为负和正的变化情况。

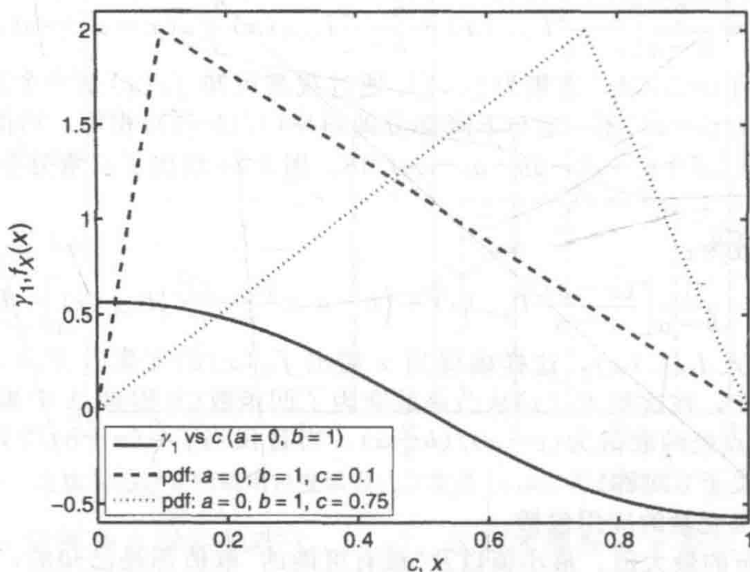


图 3-25 例 3-30 中，三角分布的斜率系数和概率密度函数

3.8.9 瑞利随机变量

瑞利分布随机变量的概率密度函数为：

$$f_X(x) = (x/\alpha^2)\exp(-x^2/2\alpha^2)I_{[0,\infty)}(x) \tag{3-126}$$

其中 $\alpha>0$ 为比例因子。累积分布函数为：

$$F_X(x) = [1-\exp(-x^2/2\alpha^2)]I_{[0,\infty)}(x) \tag{3-127}$$

图 3-26 给出了 α 取不同值时瑞利分布概率密度函数的一些例子。均值为 $\mu_X=\alpha\sqrt{\pi/2}$，方差为 $\sigma_X^2=(4-\pi)\alpha^2/2$。对高斯随机变量进行求和时就会用到瑞利随机变量。特别的，如果 X_1 和 X_2 是独立的高斯随机变量，分布参数都为 $\{\mu=0,\sigma\}$，那么 Y 就是参数为 $\alpha=\sigma$ 的瑞利随机变量。

$$Y=\sqrt{X_1^2+X_2^2} \tag{3-128}$$

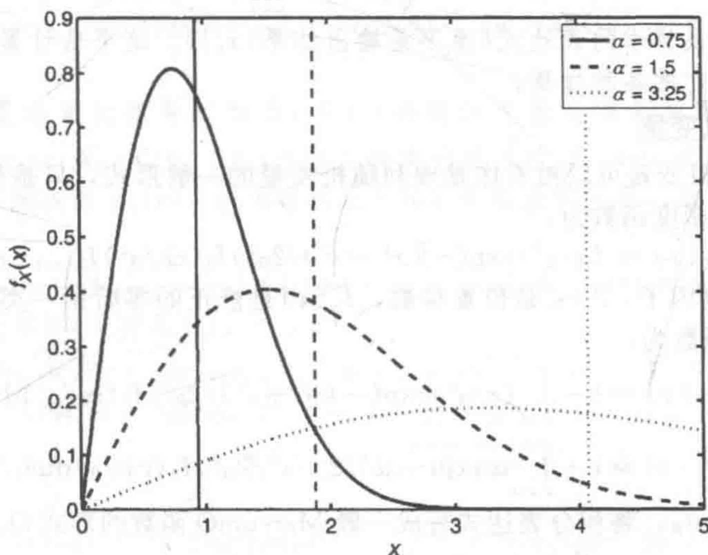


图 3-26 瑞利分布概率密度函数的一些例子，支集是$\mathcal{R}^+$，垂线标出了 μ_X 的位置

这个结果可以利用第 4 章介绍的变换方法推导得到。

瑞利分布的应用场合包括：

- 在无线通信信道中对一个信号幅度变化进行建模(称作衰落)，假设没有明显的视线(LoS)传播。
- 超声图像中的斑点噪声。
- 对风速变化进行建模。

例 3-31 计算 $F_X(0.5)$，得到瑞利随机变量的中位数为：

$$\exp(-m_e^2/2\alpha^2)=1/2\Rightarrow m_e=\alpha\sqrt{2\ln(2)} \tag{3-129}$$ ◀

例 3-32 因为瑞利分布的概率密度函数是单峰的，它的众数可以通过对 $f_X(x)$关于 x 求导来计算，将求导结果设为 0：

$$(1/\alpha^2)\exp(-m_o^2/2\alpha^2)-(m_o/\alpha^2)\exp(-m_o^2/2\alpha^2)(m_o/\alpha^2)=0 \tag{3-130}$$

消掉多余项，求解 m_o 可得：

$$1-(m_o/\alpha)^2=0\Rightarrow m_o=\alpha \tag{3-131}$$

因此，瑞利分布的均值、中位数和众数都与比例因子 α 成比例。如果分布的宽度增加，这些值都会往右边平移。 ◀

例 3-33 在第 9 章中，我们将要学习如何根据 N 个独立同分布(iid)的样本 $\{X_1,\cdots,X_N\}$ 来估计一个分布的参数。在本例中，我们简要介绍一下最大似然(ML)估计法，参数

θ 可由如下方法估计：

$$\theta_{ML} = \arg\max_{\theta} \prod_{n=1}^{N} f_X(x_n) \tag{3-132}$$

$\{X_n\}$ 的联合分布是乘积的形式，这是因为它们都是独立同分布的(联合分布将在第 4 章介绍)。因为 $\ln(\cdot)$ 是一个单调函数，计算最大似然估计的一个简便方法就是对等式右边取对数最大值。对于瑞利分布，$\theta=\alpha$，因此有：

$$\ln\left(\prod_{n=1}^{N} f_X(x_n)\right) = -N\ln(\alpha^2) + \ln\left(\prod_{n=1}^{N} x_n\right) - (1/2\alpha^2)\sum_{n=1}^{N} x_n^2 \tag{3-133}$$

对式(3-133)关于 α 求导，并令求导结果为 0，可计算出 α 的最大似然估计值为：

$$\alpha_{ML} = \sqrt{(1/2N)\sum_{n=1}^{N} X_n^2} \tag{3-134}$$

在此我们将结果写成样本的表达式(而不是输出结果 $\{x_n\}$)。这个估计器是一个随机变量，在第 9 章将要介绍它的各种性质。◀

3.8.10 莱斯随机变量

莱斯(Rice)随机变量可以被看作是瑞利随机变量的一般形式，只是位置有所变化。莱斯随机变量的概率密度函数为：

$$f_X(x) = (x/\alpha^2)\exp(-(x^2+c^2)/2\alpha^2)I_0(cx/\alpha^2)I_{[0,\infty)}(x) \tag{3-135}$$

其中 $\alpha>0$ 表示比例因子，$c>0$ 是位置参数，$I_0(x)$ 是修正的零阶第一类贝赛尔函数(见附录 B)。累积分布函数为：

$$\begin{aligned} F_X(x) &= 1 - \int_x^{\infty} (v/\alpha^2)\exp(-(v^2+c^2)/2\alpha^2)I_0(cv/\alpha^2)\mathrm{d}v \\ &= 1 - \int_{x/\alpha}^{\infty} w\exp(-w^2/2 - c^2/2\alpha^2)I_0(cw/\alpha)\mathrm{d}w \end{aligned} \tag{3-136}$$

进行变量替换 $w=v/\alpha$，将积分表达式写成一般 MarcumQ 函数的形式 $Q_m(a, b)$，可得：

$$F_X(x) = 1 - Q_1(c/\alpha, x/\alpha)I_{[0,\infty)}(x) \tag{3-137}$$

莱斯分布概率密度函数的例子如图 3-27 所示。当 $c=0$ 时莱斯随机变量退化为瑞利随机变量，此时 $I_0(cx/\alpha^2)=1$。均值为 $\mu_X=\alpha\sqrt{\pi/2}L_{1/2}(-c^2/2\alpha^2)$，方差为 $\sigma_X^2=c^2+2\alpha^2-(\pi\alpha^2/2)L_{1/2}^2(-c^2/2\alpha^2)$，其中 $L_n(x)$ 是拉盖尔多项式，也可以写成合流超几何函数的形式(见附录 B)。

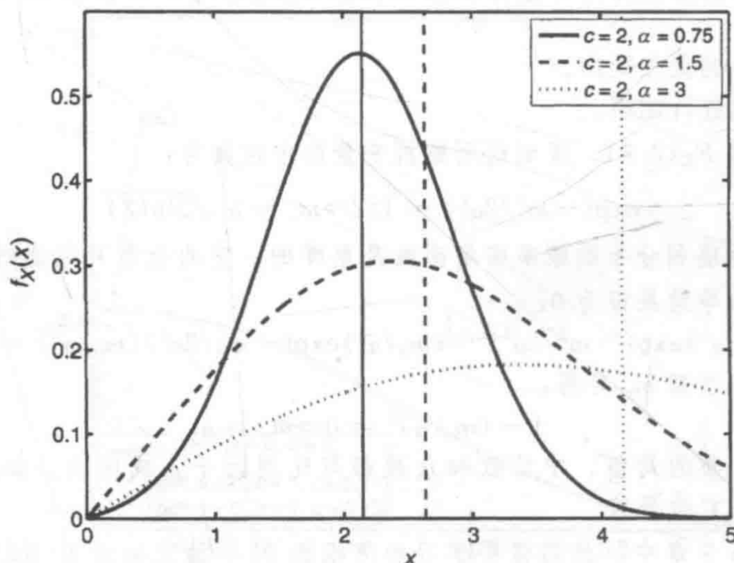


图 3-27 莱斯概率密度函数的例子，支集为 $\mathcal{R}^+$，垂线指示了 μ_X 的位置

莱斯随机变量的应用场合包括：

- 在无线通信信道中没有明显的视线传播时(从发送端直接到接收端)，接收信号幅度的变化。
- 磁共振成像中的噪声模型。

与瑞利随机变量类似，莱斯随机变量也可以通过两个独立的高斯随机变量来产生。根据式(3-128)，高斯随机变量的标准差为$\sigma=\alpha$，但是$\{X_1, X_2\}$不能取$c\cos(\theta)$和$c\sin(\theta)$形式的非零均值，$\theta\in[-\pi, \pi]$。X_1的均值可以取任何值，X_2的均值也是与其有联系的，因为式(3-128)对角度是不变的，并且$c^2=c^2\cos^2\theta+c^2\sin^2\theta$。概率密度函数的推导请见例3-34。

例3-34 设X_1和X_2是独立的高斯随机变量，分布参数分别为$\{\mu_1=c\cos(\theta),\sigma\}$和$\{\mu_2=c\sin(\theta),\sigma\}$。它们的联合概率密度函数(定义见第4章)就是独立(边缘)概率密度函数的简单乘积：

$$f_{X_1,X_2}(x_1,x_2)=(1/2\pi\sigma^2)\exp(-[x_1-c\cos(\theta)]^2/2\sigma^2)\exp(-[x_2-c\sin(\theta)]^2/2\sigma^2) \tag{3-138}$$

莱斯概率密度函数的推导过程为：(ⅰ)将随机变量变换到极坐标系下，即$R=\sqrt{X_1^2+X_2^2}$，$\Phi=\arctan(X_2/X_1)$。(ⅱ)对联合概率密度函数$f_{R,\Phi}(r,\phi)$关于ϕ积分，得到莱斯分布的概率密度函数$f_R(r)$，因为根据定义$\mathcal{R}$是莱斯随机变量。对于极坐标系，有下面的结论：

$$x_1=r\cos(\phi),x_2=r\sin(\phi)\Rightarrow r=\sqrt{x_1^2+x_2^2},\phi=\arctan(x_2/x_1) \tag{3-139}$$

得到下面的雅可比矩阵(见附录G)：

$$\boldsymbol{J}=\begin{pmatrix}\partial r/\partial x_1 & \partial r/\partial x_2\\ \partial\phi/\partial x_1 & \partial\phi/\partial x_2\end{pmatrix}=\begin{bmatrix}x_1/\sqrt{x_1^2+x_2^2} & x_2/\sqrt{x_1^2+x_2^2}\\ -x_2/(x_1^2+x_2^2) & x_1/(x_1^2+x_2^2)\end{bmatrix} \tag{3-140}$$

雅可比值J是$\boldsymbol{J}$的行列式，可得$|J|=1/\sqrt{x_1^2+x_2^2}=1/r$。因此，$R$和$\Phi$的联合概率密度函数为：

$$f_{R,\Phi}(r,\phi)=(r/2\pi\sigma^2)\exp(-[r^2+c^2-2rc\cos(\phi-\theta)]/2\sigma^2) \tag{3-141}$$

其中上式用到了$\cos(\phi-\theta)=\cos(\theta)\cos(\phi)+\sin(\theta)\sin(\phi)$。对$\phi$进行积分可得：

$$\begin{aligned}f_R(r)&=(r/2\pi\sigma^2)\exp(-[r^2+c^2]/2\sigma^2)\int_0^{2\pi}\exp([rc/\sigma^2]\cos(\phi-\theta))\mathrm{d}\phi\\&=(r/\sigma^2)\exp(-[r^2+c^2]/2\sigma^2)I_0(rc/\sigma^2)\end{aligned} \tag{3-142}$$

我们在此用到了下面的结论，即cos函数在一个完整区间上的积分可以写成下面的形式：

$$(1/2\pi)\int_0^{2\pi}\exp([rc/\sigma^2]\cos(\phi-\theta))\mathrm{d}\phi=(1/2\pi)\int_0^{2\pi}\exp([rc/\sigma^2]\cos(\phi))\mathrm{d}\phi=I_0(rc/\sigma^2) \tag{3-143}$$

上式对于任意的相移θ都成立。式(3-142)的结果就是莱斯概率密度函数(见式3-135，$\sigma=\alpha$，$r=x$)。 ◀

3.8.11 伽马随机变量($r\in\mathcal{N}$时为埃尔朗(Erlang)分布)

伽马(Gamma)随机变量的概率密度函数为：

$$f_X(x)=\frac{1}{\Gamma(r)}x^{r-1}\lambda^r\exp(-\lambda x)I_{[0,\infty)}(x) \tag{3-144}$$

其中$\lambda>0$为比例因子，$r>0$为形状参数，$\Gamma(r)$是伽马函数(见附录B)。伽马概率密度函数的例子见图3-28。累积分布函数无法写成闭合表达式，为了方便一般可以写成如下表达式：

$$F_X(x)=\frac{\gamma(r,\lambda x)}{\Gamma(r)}I_{[0,\infty)}(x) \tag{3-145}$$

其中下式是下不完全伽马函数：

$$\gamma(r,\lambda x)\triangleq\int_0^{\lambda x}v^{r-1}\exp(-v)\mathrm{d}v \tag{3-146}$$

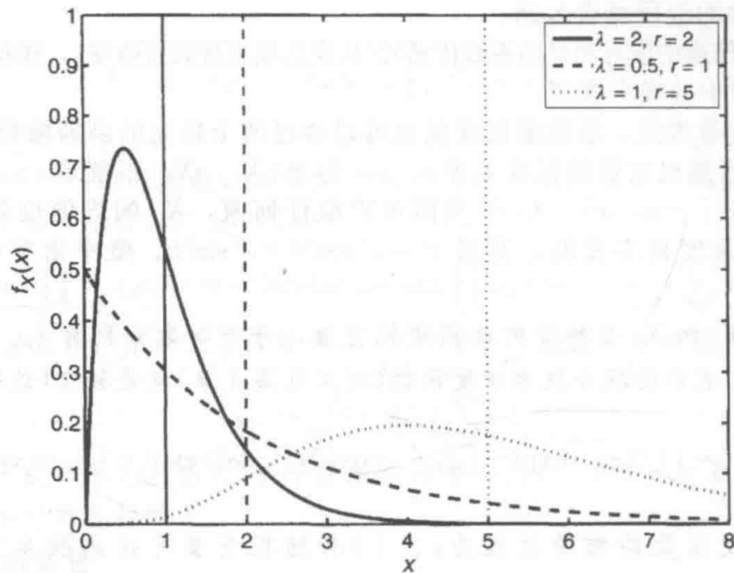


图 3-28 伽马概率密度函数的例子，支集为$\mathcal{R}^+$，垂线指示了 μ_X 的位置

对于任意 $r\in\mathcal{R}$，伽马函数 $\Gamma(r)$由 $\gamma(r,\infty)$ 给出。因此，"不完全"的意思是 $\gamma(r,\lambda x)$ 的积分是在有限值 $\lambda x<\infty$上进行的。我们可以看出，下不完全伽马函数的作用与高斯累积分布函数中的 Q 函数类似，同样也没有闭合表达式。

表 3-2 伽马随机变量的特例

伽马参数	分布类型
$r=1$	指数$\{\lambda\}$
$r\in\mathcal{N}$	埃尔朗$\{\lambda,r\}$
$r=N$，$\lambda=1/2$	卡方$\{N\}$

当分布参数取某些特定值时，伽马分布函数退化为其他类型的分布，表 3-2 对这些情况进行了总结。指数分布随机变量是一种一般情况，这是对第 r 次到达时间建模。当 $r\in\mathcal{N}$时，伽马函数简化为：

$$\Gamma(r)\triangleq(r-1)! \tag{3-147}$$

此时伽马随机变量服从埃尔朗分布。通过变换$Y=\sqrt{X}$(见习题 4-29)，伽马分布退化为 Nakagami 分布(Proakis，2001)，Nakagami 分布是无线通信信道中的一种衰落模型(另外两种衰落信道模型为瑞利分布和 Rice 分布)。Nakagami 分布的概率密度函数为：

$$f_Y(y)=\frac{2r^r}{\Gamma(r)\alpha^r}y^{2r-1}\exp(-ry^2/\alpha)I_{[0,\infty)}(y) \tag{3-148}$$

其中 $r\geqslant1/2$ 和 $\alpha>0$ 分别是形状参数和比例因子。Y 的均值为 $\mu_Y=\sqrt{r/\alpha}\,\Gamma(r+1/2)/\Gamma(1/2)$，方差为 $\sigma_Y^2=\alpha[1-(1/r)\Gamma^2(r+1/2)/\Gamma^2(r)]$，累积分布函数为：

$$F_X(y)=\gamma(r,ry^2/\alpha)/\Gamma(r) \tag{3-149}$$

伽马分布的应用场合包括：

- 泊松过程中等待第 r 个事件发生时的等待时间(见第 6 章)。
- 财经事务中的风险分析，比如对保险索赔额度的建模。
- 伽马过程是基于伽马分布的一个独立增量过程(见第 6 章)，常用于对降雨量的建模。

例 3-35 尽管中位数是分布参数的隐函数，但相对而言众数还是比较容易计算的。对概率密度函数求导，将求导结果设为 0，求解 m_o 可得($r>1$)：

$$\frac{1}{\Gamma(r)}(r-1)m_o^{r-2}\exp(-\lambda m_o)-\frac{1}{\Gamma(r)}m_o^{r-1}\exp(-\lambda m_o)\lambda=0\Rightarrow m_o=(r-1)/\lambda \tag{3-150}$$

通过计算可知当 $r\in(0,1]$ 时 $m_o=0$。◀

例 3-36 当 $\lambda=1$ 时为标准伽马概率密度函数：

$$f_X(x)=\frac{1}{\Gamma(r)}x^{r-1}\exp(-x)I_{[0,\infty)}(x) \tag{3-151}$$

通过例 3-35 可知，众数 $m_o=r-1$。对 x 进行两次求导可得：

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}f_X(x)=\frac{1}{\Gamma(r)}x^{r-3}\exp(-x)[x^2-2(r-1)x+(r-1)(r-2)] \tag{3-152}$$

将上式设为 0 再求解 x，可知标准伽马概率密度函数存在以下两个拐点：

$$x=m_o\pm\sqrt{m_o} \tag{3-153}$$

从 m_o 开始，当 x 从拐点的左边移到右边时，概率密度函数从凹函数变为了凸函数。从图 3-28可以看出，当 $r=5$，$\lambda=1$ 时，众数为 $m_o=4$，该曲线的两个拐点分别位于 $x=\{2,6\}$。◀

例 3-37 伽马分布的均值计算过程如下：

$$\mu_X=\frac{\lambda^r}{\Gamma(r)}\int_0^\infty x^r\exp(-\lambda x)\mathrm{d}x=\frac{\lambda^{r-1}}{\Gamma(r)}\int_0^\infty \lambda x^r\exp(-\lambda x)\mathrm{d}x \tag{3-154}$$

上面积分为指数随机变量的第 r 阶原始矩，根据附录 A 可知积分结果为 $\Gamma(r+1)/\lambda^r$。又因为 $\Gamma(r+1)=r\Gamma(r)$，所以可得：

$$\mu_X=\Gamma(r+1)/\lambda\Gamma(r)=r/\lambda \tag{3-155}$$

可知伽马分布均值为指数分布均值的 r 倍。同样的，方差 $\sigma_X^2=r/\lambda^2$ 也是 r 倍。◀

3.8.12 贝塔随机变量($\alpha=\beta=1/2$ 时为反正弦，$\beta=1$ 时为幂函数)

贝塔(Beta)随机变量的概率密度函数为：

$$f_X(x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}I_{[0,1]}(x) \tag{3-156}$$

其中

$$B(\alpha,\beta)\triangleq\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{3-157}$$

α，$\beta>0$ 为形状参数，$\Gamma(a)$为伽马函数。除了均匀分布和三角随机变量外，贝塔分布是本书介绍的最后一个具有有限支集的连续随机变量。贝塔分布的累积分布函数为：

$$F_X(x)=\frac{B_x(\alpha,\beta)}{B(\alpha,\beta)}I_{[0,1]}(x)+I_{(1,\infty)}(x) \tag{3-158}$$

其中下式为不完全贝塔函数。

$$B_x(\alpha,\beta)\triangleq\int_0^x y^{\alpha-1}(1-y)^{\beta-1}\mathrm{d}y \tag{3-159}$$

当 $B_x(\alpha,\beta)$ 定积分的上限 $x=1$ 时，对应的就是完全贝塔函数 $B(\alpha,\beta)$。累积分布函数中的比值称作规则不完全贝塔函数(附录 B 详细谈论了这些函数)：

$$I_x(\alpha,\beta)\triangleq\frac{B_x(\alpha,\beta)}{B(\alpha,\beta)} \tag{3-160}$$

贝塔分布的均值为 $\mu_X=\alpha/(\alpha+\beta)$，方差为 $\sigma_X^2=\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$。

Beta 随机变量的应用场合包括：

- 对样本(测量值)的基本分布进行建模，并且样本的取值被限制在一个有限范围内。
- 大气中太阳辐射的透射比。
- 对完成一个工程所需工作时间的建模，在 PERT(Project Evaluation and Review Technique，项目评估法)中经常用到(de Neufville and Stafford，1971)。

如果需要将贝塔随机变量的取值范围从 $[0,1]$ 扩展到 $[a,b]$，可以采用下面更一般的概率密度函数表达式：

$$f_X(x)=\frac{1}{B(\alpha,\beta)(b-a)^{\alpha+\beta-1}}(x-a)^{\alpha-1}(b-x)^{\beta-1}I_{[a,b]}(x) \tag{3-161}$$

贝塔概率密度函数的外形与其分布参数有很大关系。表 3-3 归纳了特例，图 3-29 给出了贝塔概率密度函数的一些例子。可以看出，概率密度函数的外形只与 $x^{\alpha-1}(1-x)^{\beta-1}$ 这部分有关系，而 $1/B(\alpha,\beta)$ 只是一个权重，要保证 $f_X(x)$ 下方的面积为 1。(附录 D 中详细归纳了连续随机变量的函数表达式。)如果 $\alpha=\beta=1$，此时 $B(1,1)=1$，式(3-156)为：

$$f_X(x)=I_{[0,1]}(x) \tag{3-162}$$

表 3-3　贝塔随机变量的外形和一些特例

贝塔参数	外形
$\alpha=1$，$\beta>2$(或 $\beta=1$，$\alpha>2$)	严格凸
$\alpha=1$，$1<\beta<2$(或 $\beta=1$，$1<\alpha<2$)	严格凹
$\alpha=1$，$\beta=2$(或 $\alpha=2$，$\beta=1$)	直线
$\alpha=\beta$	关于 1/2 对称
特例	**分布**
$\alpha=\beta=1/2$	反正弦
$\alpha=\beta=1$	[0，1]区间上的标准均匀分布
$\beta=1$	幂函数{α}

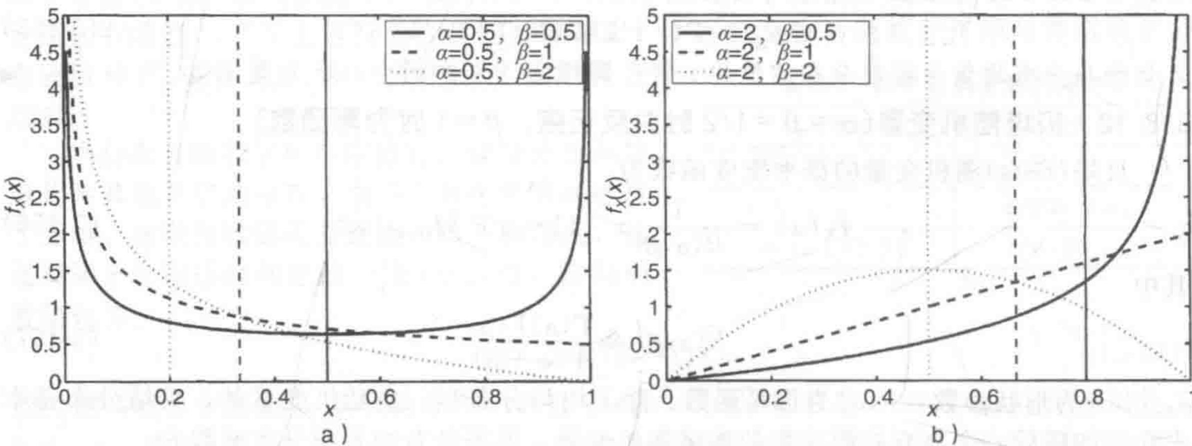


图 3-29　贝塔概率密度函数的一些例子，支集是[0,1]，垂线指示了 μ_X 的位置。a) $\alpha=0.5$；b) $\alpha=2$

上式是[0,1]区间连续均匀分布随机变量的概率密度函数。当 $\alpha=\beta=1/2$ 时，得到反正弦分布：

$$F_X(x)=(2/\pi)\arcsin(\sqrt{x})I_{[0,1]}(x)+I_{(1,\infty)}(x)\Rightarrow f_X(x)=1/\pi\sqrt{x(1-x)}I_{[0,1]}(x) \tag{3-163}$$

当 $\beta=1$ 时，代入 $B(\alpha,1)=\alpha$，得到幂函数分布的表达式：

$$f_X(x)=\alpha x^{\alpha-1}I_{[0,1]}(x) \tag{3-164}$$

$\alpha=r_1$，$\beta=r_2$ 的贝塔分布可以通过下面的变换计算：

$$X=\frac{Y_1}{Y_1+Y_2} \tag{3-165}$$

其中 Y_1 和 Y_2 分别是参数为 $\{r_1,\lambda\}$ 和 $\{r_2,\lambda\}$ 的伽马随机变量。很显然，X 的分布关于 Y_1 和 Y_2 都不是对称的。

例 3-38　贝塔分布的斜率为：

$$\gamma_1=\sqrt{\frac{\alpha+\beta+1}{\alpha\beta}}\frac{2(\beta-\alpha)}{(\alpha+\beta+2)} \tag{3-166}$$

从上式可以看出，分子中的 $\beta-\alpha$ 决定了斜率的类型。图 3-29a 给出了斜率为正的两个例子

($\beta>\alpha$)，第三个例子是对称的($\alpha=\beta=1/2$)。图 3-29b 中的概率密度函数斜率为负($\beta<\alpha$)，另外一种也是对称的($\alpha=\beta=1/2$)。◀

例 3-39 接下来，我们证明指数分布可以通过变换 $Y=\beta X$ 得到，其中 X 是贝塔随机变量，分布参数为 $\{\alpha=1,\beta\}$，设 $\beta\to\infty$。Y 的累积分布函数为：

$$P(Y\leqslant y)=P(\beta X\leqslant y)=P(X\leqslant y/\beta) = \frac{B_{y/\beta}(1,\beta)}{B(1,\beta)}=\frac{1}{B(1,\beta)}\int_0^{y/\beta}(1-v)^{\beta-1}\mathrm{d}v \tag{3-167}$$

当 β 与 X 相乘后支集为 $[0,\beta)$。因此可得：

$$F_Y(y)=-\frac{1}{B(1,\beta)}(1/\beta)(1-v)^{\beta}\big|_0^{y/\beta}=\frac{\Gamma(1+\beta)}{\Gamma(1)\Gamma(\beta)}(1/\beta)[1-(1-y/\beta)^{\beta}] \tag{3-168}$$

又因为 $\Gamma(1)=1$，$\Gamma(1+\beta)/\Gamma(\beta)=\beta$，最后可得：

$$\lim_{\beta\to\infty}F_Y(y)=[1-\exp(-y)]I_{[0,\infty)}(x) \tag{3-169}$$

上式为标准指数分布($\lambda=1$)的累积分布函数。类似的，如果 X 的分布参数为 $\{\alpha=r,\beta\}$，那么 Y 的极限分布为伽马分布，分布参数为 $\{\lambda=1,r\}$（见习题 3-22）。◀

3.8.13 帕雷托随机变量

帕雷托(Pareto)随机变量的概率密度函数为：

$$f_X(x)=(r\alpha^r/x^{r+1})I_{[\alpha,\infty)}(x) \tag{3-170}$$

其中 $\alpha>0$ 为比例因子，并通过指示函数定义了支集。$r>0$ 是形状参数。对于非常小的 r，分布函数也存在非常严重的右拖尾现象。累积分布函数为：

$$F_X(x)=[1-(\alpha/x)^r]I_{[\alpha,\infty)}(x) \tag{3-171}$$

帕雷托概率密度函数的例子如图 3-30 所示。均值为 $\mu_X=r\alpha/(r-1)$，$r>1$。方差为 $\sigma_X^2=r\alpha^2/[(r-1)^2(r-2)]$，$\alpha>2$(否则，均值和方差没有定义)。

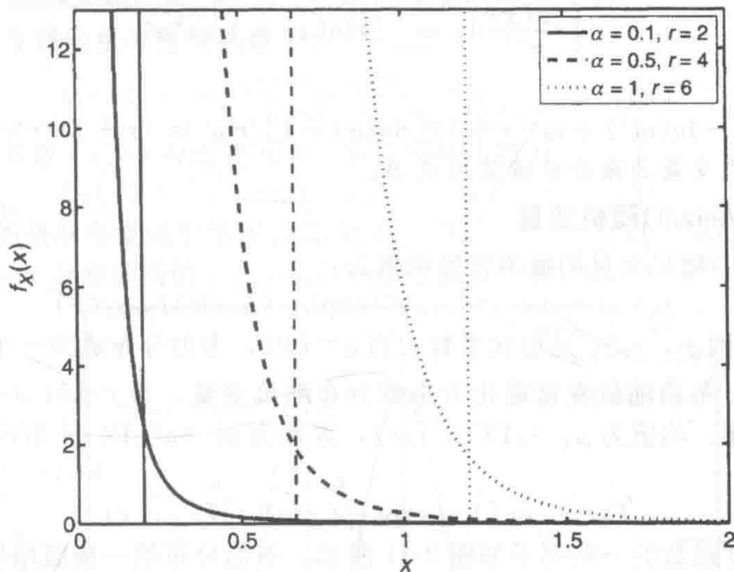


图 3-30 帕雷托概率密度函数的例子，支集为 $[\alpha,\infty)$，垂线指示了 μ_X 的位置

帕雷托随机变量的应用场合包括：

- 收入和健康的分布。
- 通信电路和硬盘驱动器中的错误簇。

例 3-40 对于较大的 r 取值，可以用帕雷托概率密度函数来近似狄拉克 δ 函数：

$$\lim_{r\to\infty}f_X(x)=\delta(x-\alpha) \tag{3-172}$$

那么，帕雷托累积分布函数就可以近似单位阶跃函数：

$$\lim_{r\to\infty} F_X(x) = u(x-\alpha) \tag{3-173}$$

首先，我们验证一下概率密度函数的积分是否为1：

$$r\alpha^r \int_\alpha^\infty x^{-r-1}\mathrm{d}x = -\alpha x^{-r}\Big|_\alpha^\infty = 1 \tag{3-174}$$

然后，通过观察可得：

$$\lim_{r\to\infty}[r\alpha^r/x^{r+1}u(x-\alpha)] = (1/x)u(x-\alpha)\lim_{r\to\infty} r(\alpha/x)^r = \begin{cases}0, & x>\alpha\\ \infty, & x=\alpha\end{cases} \tag{3-175}$$

其中 $u(x)$ 就是单位阶跃函数，并且根据定义 $\alpha/x\leqslant 1$。式(3-174)和(3-175)共同给出了狄拉克 δ 函数在 $x=\alpha$ 处的定义。在定义狄拉克 δ 函数时，并不需要原始函数一定是偶函数，在附录B中使用的是矩形函数。◀

例 3-41 (微分熵) **微分熵**是一个连续随机变量，是信息论中经常用到的一个量。它的定义如下所示(见第10章)：

$$h(X) \triangleq -\int_{-\infty}^{\infty} f_X(x)\log(f_X(x))\mathrm{d}x \tag{3-176}$$

式(3-176)的对数函数可以任何数为底，不过我们经常用2或 e 作底，那么 $h(X)$ 的单位就是比特(bits)或莱特(nats)。对于帕雷托随机变量：

$$\begin{aligned} h(X) &= -\int_\alpha^\infty \frac{r\alpha^r}{x^{r+1}}\ln(r\alpha^r)\mathrm{d}x + \int_\alpha^\infty \frac{r\alpha^r}{x^{r+1}}\ln(x^{r+1})\mathrm{d}x \\ &= -\ln(r\alpha^r) + r\alpha^r(r+1)\int_\alpha^\infty \frac{\ln(x)}{x^{r+1}}\mathrm{d}x \end{aligned} \tag{3-177}$$

因为对数运算可以从第一项积分表达式中提取出来，因此其运算结果可以简化。剩下的部分是对帕雷托概率密度函数进行积分。第二项的不定积分结果为：

$$\int \frac{\ln(x)}{x^{r+1}}\mathrm{d}x = -[r\ln(x)+1]/r^2x^r \tag{3-178}$$

因此可得：

$$h(X) = -\ln(r\alpha^r) + r\alpha^r(r+1)[r\ln(\alpha)+1]/r^2\alpha^r = (r+1)/r + \ln(\alpha/r) \tag{3-179}$$

其他类型随机变量的微分熵请见附录A。◀

3.8.14 韦伯(Weibull)随机变量

韦伯(Weibull)随机变量的概率密度函数为：

$$f_X(x) = (r/\alpha)(x/\alpha)^{r-1}\exp(-(x/\alpha)^r)I_{[0,\infty)}(x) \tag{3-180}$$

其中 $\alpha>0$ 为比例因子，$r>0$ 是形状参数。当 $\alpha=1$ 时，韦伯分布成为一个单参数函数。当 $r=1$，$\alpha=1/\lambda$ 时，韦伯随机变量退化为指数分布随机变量。当 $r=2$，$\alpha\to\sqrt{2}\alpha$ 时，韦伯分布退化为瑞利分布。均值为 $\mu_X=\alpha\Gamma(1+1/r)$，方差为 $\sigma_X^2=\alpha^2[\Gamma(1+2/r)-\Gamma^2(1+1/r)]$，累积分布函数为：

$$F_X(x) = [1-\exp(-(x/\alpha)^r)]I_{[0,\infty)}(x) \tag{3-181}$$

韦伯概率密度函数的一些例子如图3-31所示。韦伯分布的一些应用场合包括：

- 风速的分布，以及风力涡轮机阵的功率模型。
- 复杂系统中性能损失和器件置信度的模型。
- 无线通信系统中的衰落信道模型。

例 3-42 可知对于任意 r，$F_X(\alpha)=1-\exp(-1)\approx 0.6321$。因此在图3-31中，函数在 $x=\alpha$ 以左下方的面积都是相等的，即使随着 r 的增大，概率密度函数变得越来越扁，该结论也是成立的。对 $f_X(x)$ 关于 x 求导，并将求导结果设为0，可以计算出众数为($r>1$)：

$$m_o = \alpha[(r-1)/r]^{1/r} \tag{3-182}$$

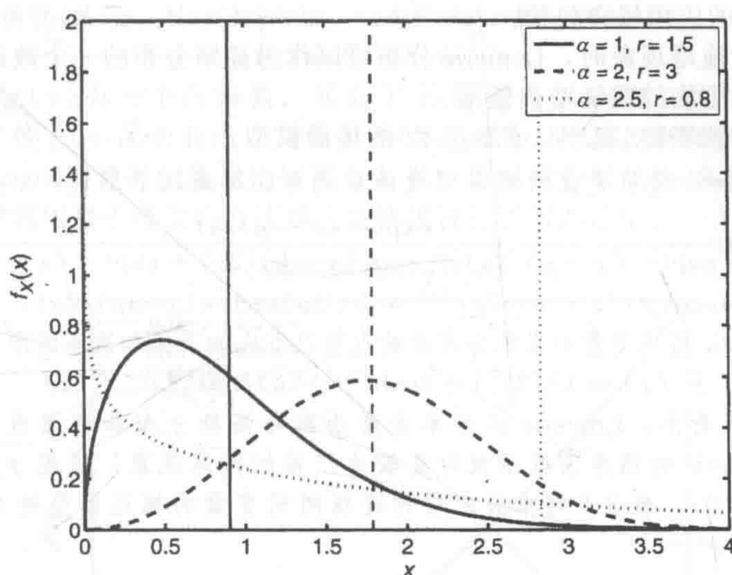


图 3-31　韦伯概率密度函数的例子，支集为$\mathcal{R}^+$，垂线指示了 μ_X 的位置

当 $r=2$ 时众数为 0.7071α，$r=5$ 时众数为 0.9564α。当 $r\in(0,1]$ 时，众数为 0。中位数的计算过程如下：

$$1-\exp(-(m_e/\alpha)^r)=1/2\Rightarrow m_e=\alpha(\ln(2))^{1/r} \tag{3-183}$$

当 $r=2$ 时中位数为 0.8362α，$r=5$ 时中位数为 0.9293α。比较 m_o 和 m_e 的表达式可知，当 $r>1/(1-\ln(2))\approx3.2589$ 时，$m_o>m_e$。众数和中位数不可能超过 $x=\alpha$，因为累积分布函数在该处的取值为 0.6321。◀

3.8.15 Logistic 随机变量（{μ=0，α=1}时为 Sigmoid 分布）

Logistic 随机变量的概率密度函数为：

$$f_X(x)=\frac{\exp(-(x-\mu)/\alpha)}{\alpha(1+\exp(-(x-\mu)/\alpha))^2}I_{(-\infty,\infty)}(x) \tag{3-184}$$

其中 $\mu\in\mathcal{R}$为位置参数，$\alpha>0$ 为比例因子。累积分布函数为：

$$F_X(x)=[1+\exp(-(x-\mu)/\alpha)]^{-1}I_{(-\infty,\infty)}(x) \tag{3-185}$$

Logistic 分布的概率密度函数如图 3-32 所示。当 $\mu=0$，$\alpha=1$ 时，Sigmoid 分布为$F_X(x)$的一个特例。Logistic 分布的均值为 $\mu_X=\mu$，方差为 $\sigma_X^2=(\pi\alpha)^2/3$。

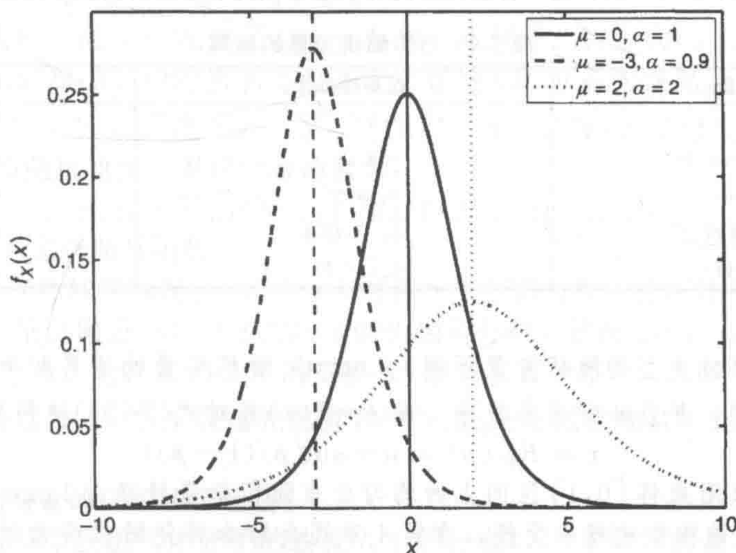


图 3-32　Logistic 分布的概率密度函数，支集为$\mathcal{R}$，垂线指示了 μ_X 的位置

Logistic 分布的应用场合包括：

- 当存在轻微拖尾现象时，Logistic 分布可以作为高斯分布的一个改进模型。
- 人类或其他生物的数量增长模型。
- 人类社会中新事物(新产品或新技术)的传播模型。

例 3-43 Logistic 随机变量的概率密度函数还可以写成以下形式：

$$f_X(x)=\frac{\exp(-(x-\mu)/\alpha)}{\alpha[\exp(-(x-\mu)/2\alpha)]^2[\exp((x-\mu)/2\alpha)+\exp(-(x-\mu/2\alpha))]^2}$$
$$=[4\alpha\cosh^2((x-\mu)/2\alpha)]^{-1}=(1/4\alpha)\mathrm{sech}^2((x-\mu)/2\alpha) \tag{3-186}$$

同样的，Logistic 随机变量的累积分布函数也可以写成如下更加紧凑的形式(见习题 3-18)：

$$F_X(x)=(1/2)[1+\tanh((x-\mu)/2\alpha)]I_{(-\infty,\infty)}(x) \tag{3-187}$$

从图 3-33 可以看出，Logistic 的概率密度函数与高斯分布非常接近：它们的斜率 γ_1 都为 0，但是 Logistic 的概率密度函数存在较为严重的拖尾现象，峰度 $\gamma_2=6/5$(高斯概率密度函数的峰度为 0)。表 3-4 对本书中对称连续随机变量的拖尾类型进行了总结(全部类型的斜率都为 0)。

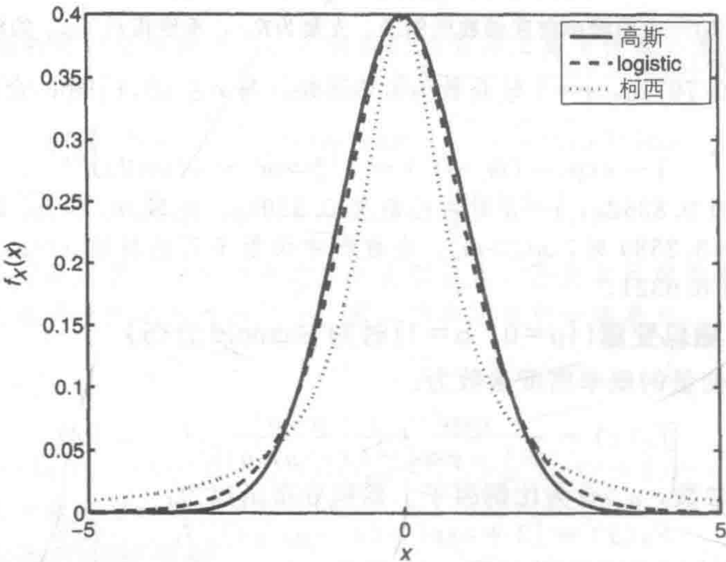


图 3-33 高斯($\sigma=1$)、Logistic($\alpha=5/8$)、柯西($\alpha=\sqrt{2/\pi}$)三种概率密度函数的对比，位置参数相同($\mu=c=0$)。选择合适的比例因子使三个概率密度函数的最大值相等

表 3-4 对称随机变量的拖尾

随机变量	过剩峰度 γ_2	拖尾类型
连续均匀	$-6/5$	轻微拖尾
三角($c=(a+b)/2$)	$-3/5$	轻微拖尾
高斯，Logistic	0，6/5	中等拖尾
拉普拉斯，柯西	3，不存在	严重拖尾
学生氏 t($r>4$)	$6/(r-4)$	严重拖尾

◀

例 3-44 与其他类型的随机变量不同，Logistic 随机变量的逆累积分布函数存在一个简单的闭合表达式，并且非常容易推导。设 $p=F(x)$并对式(3-185)进行整理后可得：

$$x=F_X^{-1}(p)=\mu+\alpha\ln(p/(1-p)) \tag{3-188}$$

这个结果可以用来将[0,1]区间上的均匀分布随机变量转换为 Logistic 分布，这在第 1 章已经讲过。这被称作逆概率变换，在第 4 章还会结合其他随机变量进行详细介绍。对于随机变量 Y，其累积分布函数为 $F_Y(y)$，均值为 μ_Y。下面这个有趣的逆变换可以用来比

较 X 和 Y 的拖尾情况(Kotz, Kozubowski, and Podgorski, 2001):

$$g(y) \triangleq F_X^{-1}(F_Y(y)) \tag{3-189}$$

若 $y \geq \mu_Y$ 时 $g(y)$ 为一个凸函数，那么 Y 的拖尾比 X 轻微。图 3-34 给出了服从 Logistic 分布的 X 的 $g(y)$ 函数曲线，其中 Y 分别是高斯、拉普拉斯和柯西分布。从图中结果可以看出，Logistic 分布的拖尾比柯西分布和拉普拉斯分布的轻微，但比高斯分布的严重，在表 3-4 中利用基于峰度的方法对这些拖尾进行了归纳总结。

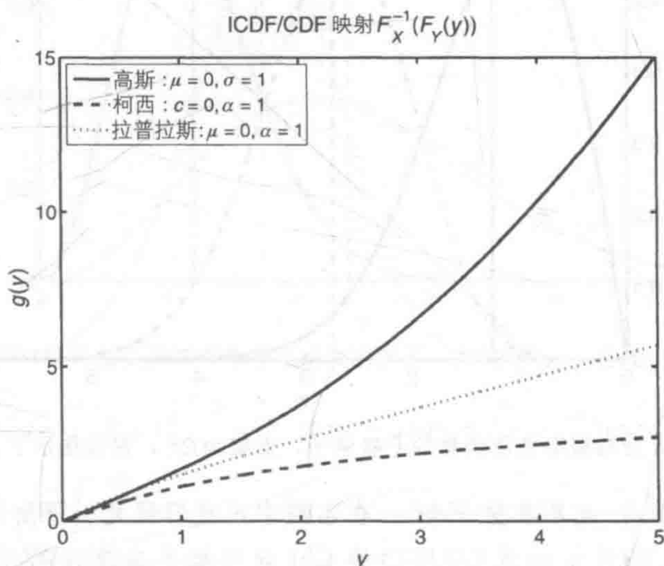


图 3-34 例 3-34 中的变换，用来判断一个分布是否有轻微的拖尾

3.8.16 Chi 分布随机变量(麦克斯韦-玻耳兹曼(Maxwell-Boltzmann)，半正态)

设$\{X_n\}$为 N 个独立的高斯随机变量，分布参数为 $\{\mu_n, \sigma_n\}$。那么下面的等式为一个 Chi 分布随机变量，

$$X \triangleq \left[\sum_{n=1}^{N}(X_n-\mu_n)^2/\sigma_n^2\right]^{1/2} \tag{3-190}$$

其概率密度函数为：

$$f_X(x) = \frac{x^{N-1}}{2^{N/2-1}\Gamma(N/2)}\exp(-x^2/2)I_{[0,\infty)}(x) \tag{3-191}$$

其中 $N \in \mathcal{N}$为自由度。图 3-35 给出了 Chi 分布的一些例子。因为式(3-190)是标准高斯随机变量之和，因此原始$\{X_n\}$的均值和方差并不会影响 Chi 分布的情况。X 的均值为 $\mu_X = \sqrt{2}\Gamma((N+1)/2)/\Gamma(N/2)$，方差为 $\sigma_X^2 = N-2[\Gamma((N+1)/2)/\Gamma(N/2)]^2$。我们经常用符号 χ_N 来表示 Chi 分布随机变量。累积分布函数为：

$$F_X(x) = [\gamma(N/2, x^2/2)/\Gamma(N/2)]I_{[0,\infty)}(x) \tag{3-192}$$

其中 $\gamma(a,b)$ 为下不完全伽马函数。

Chi 分布存在三种特殊情况：(ⅰ)$N=1$ 时称作半正态分布，此时比例因子 $\alpha=1$(也称作折叠正态分布，请见附录 A)，(ⅱ)$N=2$ 时为瑞利分布，此时 $\alpha=1$，(ⅲ)$N=3$ 时为麦克斯韦-玻耳兹曼分布，此时 $\alpha=1$。X 的变换 X^2 服从卡方分布，3.8.17 节将详细介绍它。代入 $N=2r$，进行变换$\sqrt{\alpha/2r}X$ 可得先前提到的 Nakagami 随机变量，此时 α 为比例因子，r 是形状参数。

Chi 分布随机变量的应用场合包括：

- 一个密闭容器中理想气体分子的速度大小的分布($N=3$ 对应的是麦克斯韦-玻耳兹曼分布)。

- 对缺少正负信息的高斯分布样本的建模，即 $X=|X_n-\mu_n|/\sigma_n$（$N=1$ 对应的是半正态分布）。

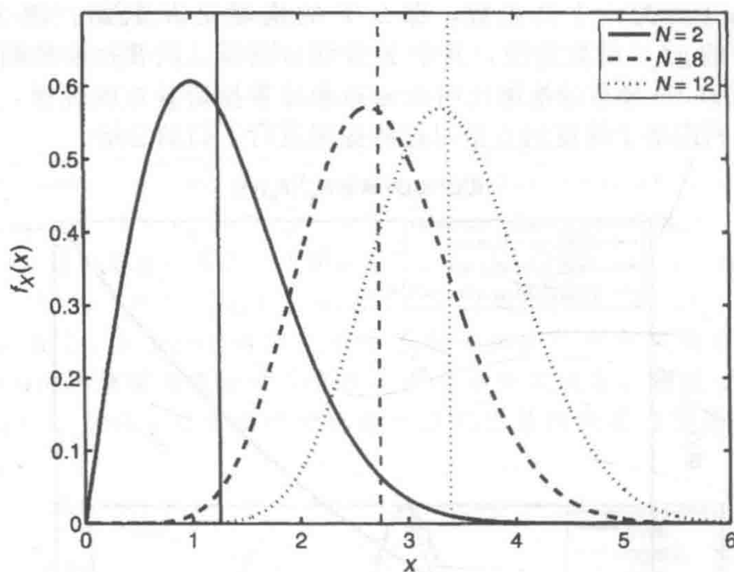


图 3-35 Chi 分布概率密度函数的一些例子，支集为$\mathcal{R}^+$，垂线指示了 μ_X 的位置

例 3-45（麦克斯韦-玻耳兹曼分布） 在本题中，我们研究一下附录 A 中的麦克斯韦-玻耳兹曼分布，比例因子为 α（式(3-181)的 Chi 分布概率密度函数没有这个参数）代入 $N=3$到 Chi 分布的概率密度函数可得：

$$f_X(x)=\sqrt{2/\pi}x^2\exp(-x^2/2)I_{[0,\infty)}(x) \tag{3-193}$$

上式已代入结论 $\Gamma(3/2)=\sqrt{\pi}/2$。假设 x^2 乘以比例因子 $1/\alpha^2$，在 X 的支集上积分可得：

$$\sqrt{2/\pi}\int_0^{\infty}(x^2/\alpha^2)\exp(-x^2/2\alpha^2)\mathrm{d}x=\alpha\sqrt{2/\pi}\int_0^{\infty}y^2\exp(-y^2/2)\mathrm{d}y=\alpha \tag{3-194}$$

对上式进行了变量替换 $y^2=x^2/\alpha^2$，并且认识到转换后的函数是一个正确的概率密度函数。乘以比例因子 α 后，麦克斯韦-玻耳兹曼分布随机变量的概率密度函数为：

$$f_X(x)=(1/\alpha^3)\sqrt{2/\pi}x^2\exp(-x^2/2\alpha^2)I_{[0,\infty)}(x) \tag{3-195}$$

在第 4 章将会介绍，上面这个结论还可以用一个更直接的方法得到。在变换 $X=\alpha Y$ 中引入了一个比例因子 $\alpha(\alpha>0)$，累积分布函数还可以写成下面的形式：

$$F_X(y)=P(X\leqslant x)=P(Y\leqslant x/\alpha)=F_Y(x/\alpha) \tag{3-196}$$

对上式进行求导可得概率密度函数：

$$f_X(x)=\frac{\mathrm{d}}{\mathrm{d}x}F_Y(x/\alpha)=(1/\alpha)f_Y(x/\alpha) \tag{3-197}$$

麦克斯韦-玻耳兹曼分布的概率密度函数如图 3-36 所示，给出了 α 不同取值的一些例子。当 $N=3$ 时，对 Chi 分布随机变量乘以比例因子 α，可以推导出均值为 $\mu_X=2\alpha\sqrt{2/\pi}$，方差为 $\sigma_X^2=\alpha^2(3\pi-8)/\pi$（见习题 3-24）。当 $N=3$ 时，Chi 随机变量的累积分布函数为：

$$F_X(x)=[1/\Gamma(3/2)][(1/2)\gamma(1/2,x^2/2)-(x/\sqrt{2})\exp(-x^2/2)]I_{[0,\infty)}(x) \tag{3-198}$$

上式用到了附录 B 中下不完全伽马函数的递归表达式，因此可得：

$$F_X(x)=(2/\sqrt{\pi})[(1/2)\int_0^{x^2/2}v^{-1/2}\exp(-v)\mathrm{d}v-(x/\sqrt{2})\exp(-x^2/2)]I_{[0,\infty)}(x) \tag{3-199}$$

为了计算麦克斯韦-玻耳兹曼分布随机变量的累积分布函数，根据式(3-196)我们引入比例因子 α 可得：

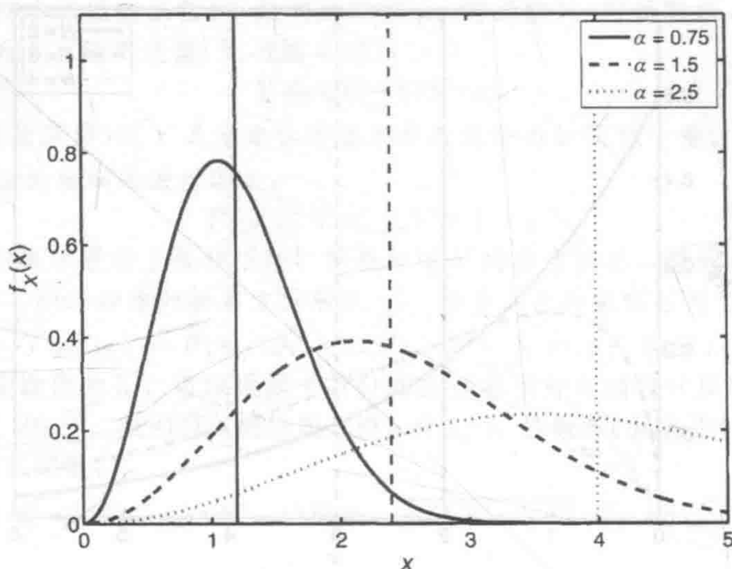


图 3-36 麦克斯韦-玻耳兹曼分布概率密度函数的例子，支集为$\mathcal{R}^+$，垂线指示了 μ_X 的位置

$$F_X(x) = (2/\sqrt{\pi})\left[(1/2)\int_0^{x^2/2\alpha^2} v^{-1/2}\exp(-v)\mathrm{d}v - (x/\alpha\sqrt{2})\exp(-x^2/2\alpha^2)\right]$$

$$= (2/\sqrt{\pi})\int_0^{x/\alpha\sqrt{2}} \exp(-w^2)\mathrm{d}w - \sqrt{2/\pi}(x/\alpha)\exp(-x^2/2\alpha^2) \quad (3\text{-}200)$$

在上式中我们进行了变量替换 $w^2=v$。因为这个积分表达式是一个误差函数，因此可以将麦克斯韦-玻耳兹曼累积分布函数写成一个更加紧凑的形式：

$$F_X(x) = [\mathrm{erf}(x/\alpha\sqrt{2}) - \sqrt{2/\pi}(x/\alpha)\exp(-x^2/2\alpha^2)]I_{[0,\infty)}(x) \quad (3\text{-}201)$$

当 $\alpha=1$ 时，上式简化为式(3-198)。◀

3.8.17 卡方分布随机变量

设$\{X_n\}$为 N 个独立的高斯随机变量，分布参数为$\{\mu_n,\sigma_n\}$。那么下式为一个卡方分布随机变量：

$$X = \sum_{n=1}^{N}(X_n-\mu_n)^2/\sigma_n^2 \quad (3\text{-}202)$$

其概率密度函数为：

$$f_X(x) = \frac{x^{N/2-1}}{2^{N/2}\Gamma(N/2)}\exp(-x/2)I_{[0,\infty)}(x) \quad (3\text{-}203)$$

其中 $N\in\mathcal{N}$表示自由度。因为式(3-202)是标准高斯随机变量之和，所以原始$\{X_n\}$的均值和方差并不会影响卡方分布的情况。我们一般采用符号 χ_N^2 来表示卡方分布随机变量。当 $N=2r$ 及 $\alpha=1/2$ 时为一个特例，此时卡方分布退化为伽马分布。当 $N=2$ 时，卡方分布随机变量退化为$\lambda=1/2$ 的指数随机变量。累积分布函数没有闭合表达式，不过可以表示成如下形式：

$$F_X(x) = \frac{1}{2^{N/2}\Gamma(N/2)}\int_0^x z^{N/2-1}\exp(-z/2)\mathrm{d}z = \frac{\gamma(N/2,x/2)}{\Gamma(N/2)}I_{(-\infty,\infty)}(x) \quad (3\text{-}204)$$

其中 $\gamma(a, b)$为下不完全伽马函数。图 3-37 给出了卡方分布概率密度函数的一些例子。均值为 $\mu_X=N$，方差为 $\sigma_X^2=2N$。通过变换可以将卡方分布转换为其他类型的分布，表 3-5 对此进行了总结。很显然，表格中的最后两个结果是根据第一行的结果得出的，因为它们是卡方分布随机变量的特例。

卡方分布的应用场合包括：

- 根据一个高斯随机变量的若干样本计算方差的置信区间。
- 假设检验来判断根据高斯样本估计的方差是否等于一个特定值。

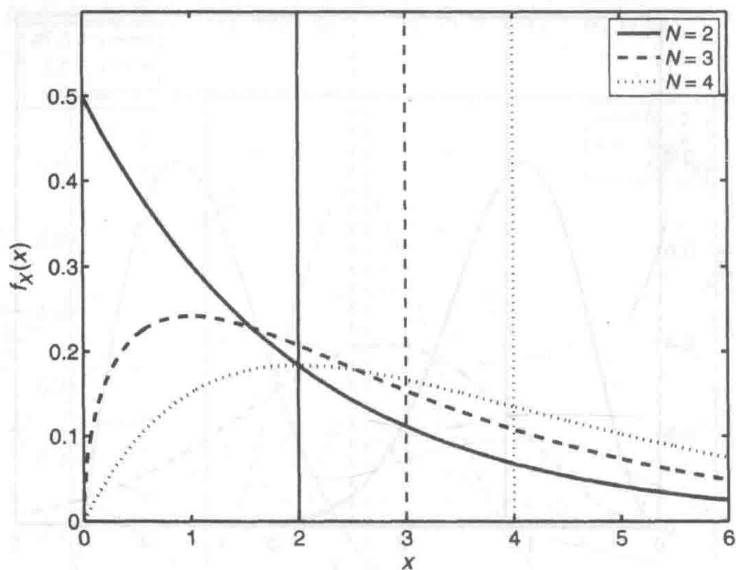


图 3-37 卡方分布概率密度函数，支集为$\mathcal{R}^+$，垂线指示了 μ_X 的位置

表 3-5 卡方分布随机变量的变换

随机变量 X	变换关系	分布
Chi 分布$\{N\}$	X^2	χ_N^2
标准高斯	X^2	χ_1^2
均匀分布[0, 1]	$-\ln(X^2)$	χ_2^2
瑞利分布$\{\alpha=1\}$	X^2	χ_2^2
麦克斯韦-玻耳兹曼$\{\alpha=1\}$	X^2	χ_3^2

例 3-46 表 3-5 中的均匀分布随机变量的变换可由 $Y=-\ln(X^2)$累积分布函数推导得到：

$$F_Y(y)=P(Y\leqslant y)=P(\ln(X^2)\geqslant -y)=1-P(X<\exp(-y/2)) \tag{3-205}$$

其中在上式我们用到了对数函数是单调函数的结论。因为 X 是区间 [0,1] 上均匀分布的随机变量，它的累积分布函数可以简化为 $xI_{[0,1]}(x)$，所以有：

$$F_Y(y)=1-\exp(-y/2)\Rightarrow f_Y(y)=(1/2)\exp(-y/2)I_{[0,\infty)}(x) \tag{3-206}$$

上式为 χ_2^2 的概率密度函数。请注意，对数运算将 X 的有限支集拓展为 Y 的$\mathcal{R}^+$。◀

例 3-47 当 $N>2$ 时，卡方分布概率密度函数的导数为：

$$\frac{\mathrm{d}}{\mathrm{d}x}f_X(x)=\frac{(N/2-1)x^{N/2-2}}{2^{N/2}\Gamma(N/2)}\exp(-x/2)-(1/2)\frac{x^{N/2-1}}{2^{N/2}\Gamma(N/2)}\exp(-x/2) \tag{3-207}$$

将求导结果设为 0，求解得到众数 $m_o=N-2$。当 $N=2$ 时：

$$\frac{\mathrm{d}}{\mathrm{d}x}f_X(x)=-(1/2)\frac{1}{2^{N/2}\Gamma(N/2)}\exp(-x/2)=0\Rightarrow m_o=0 \tag{3-208}$$

当 $N=1$ 时，将式(3-207)设为 0 并消去多余项，可以简化为 $x^{-3/2}+x^{-1/2}=0$。因为当 $N>2$ 时，剩下的两项符号肯定是相同的，所以众数 $m_o=0$。◀

在估计问题中会遇到卡方分布。设 $\{X_1,\cdots,X_N\}$ 是独立的高斯分布样本，分布参数为 $\{\mu,\sigma\}$。下面的样本均值经常被用于估计基本分布的均值：

$$\overline{X}\triangleq\frac{1}{N}\sum_{n=1}^{N}X_n \tag{3-209}$$

易知 $\overline{X}$ 服从高斯分布，均值为 μ，方差为 σ^2/N(见第 5 章)。样本方差的计算公式为：

$$S^2\triangleq\frac{1}{N-1}\sum_{n-1}^{N}(X_n-\overline{X})^2 \tag{3-210}$$

可知样本方差是样本均值的函数(不是实际均值 μ_X 的函数)。容易验证，下面的表达式是自由度为 $N-1$ 的卡方随机变量(见习题 4-35)：

$$Y \triangleq (N-1)S^2/\sigma^2 \tag{3-211}$$

例 3-48 (置信区间)设 Y 是用来估计某个参数或分布的随机变量。置信等级为 $1-\alpha_c$ 的置信区间用下面的概率来进行定义：

$$P(y_l \leqslant Y \leqslant y_u) = 1-\alpha_c \tag{3-212}$$

其中 $\{y_l, y_u\}$ 分别表示置信上限和下限。需要知道 Y 的分布情况，据此才可根据给定的 α_c 确定置信限。式(3-210)给出的样本方差服从 χ^2_{N-1} 分布，此时置信区间为：

$$P(y_l \leqslant Y \leqslant y_u) = P(y_l/(N-1)S^2 \leqslant \sigma^2 \leqslant y_u/(N-1)S^2) = 1-\alpha_c \tag{3-213}$$ ◀

对于一个特定取值的 α_c，可以根据卡方分布的逆累积分布函数计算出置信限$\{y_l, y_u\}$。通过计算样本$\{X_n\}$的 S^2，我们可以确信均方根 σ 有 $1-\alpha_c$ 的概率(也就是有 $100(1-\alpha_c)\%$的置信度)落在下面的区间里：

$$y_l/(N-1)S^2 \leqslant \sigma^2 \leqslant y_u/(N-1)S^2 \Rightarrow \sqrt{y_l/(N-1)S^2} \leqslant \sigma \leqslant \sqrt{y_u/(N-1)S^2} \tag{3-214}$$

例如，假设从标准高斯随机变量中获得了 $N=10$ 个样本，分别为 0.8369、−0.7223、−0.7215、−0.2012、−0.0205、0.2789、1.0583、0.6217、−1.7506 以及 0.6973。可计算得到样本均值 $\bar{x}=0.0077$，样本方差 $s^2=0.7687$。如果置信水平为 95%，从卡方分布的逆累积分布函数(自由度为 $N-1=9$)可以计算得到下面这些数值：$y_l=F_Y^{-1}(\alpha_c/2)=F_Y^{-1}(0.025)=2.7004$，$y_u=F_Y^{-1}(1-\alpha_c/2)=F_Y^{-1}(0.975)=19.0228$。将结论代入式(3-214)可得 95%的置信区间为：

$$0.3903 \leqslant \sigma^2 \leqslant 2.7497 \Rightarrow 0.6248 \leqslant \sigma \leqslant 1.6582 \tag{3-215}$$

而实际上，标准高斯概率密度函数的均方根为 $\sigma=1$。

例 3-49 (假设检验)在假设检验中，我们从两个或多个假设中进行选择，每个假设对应一个随机变量样本参数的取值。对于一个分布的未知方差 σ^2，假设$\{H_n\}$可能为：

$$H_0: \sigma^2 = \sigma_0^2, \quad H_1: \sigma^2 > \sigma_0^2 \tag{3-216}$$

其中 σ_0^2 为某特定值。我们再次假设我们不知道标准高斯随机变量的方差，用样本方差 S^2 来对其进行估计。对于式(3-211)中的卡方分布随机变量，显著性水平为 α_s。如果满足下面的等式，零假设 H_0 就会被否决

$$(N-1)S^2/\sigma_0^2 > F_Y^{-1}(1-\alpha_s) \tag{3-217}$$

其中 $F_Y^{-1}(\cdot)$为自由度为 $N-1$ 的卡方分布的逆累积分布函数。设 $\sigma_0^2=1.1$，$\alpha_s=0.05$。仍然使用例 3-48 中的样本，计算得到 $s^2=0.7687$。因此有：

$$(N-1)s^2/\sigma_0^2 = 9(0.7687)/1.1 = 6.2894 \tag{3-218}$$

因为在自由度为 $N-1=9$ 时，上式的结果并没有超过 $F_Y^{-1}(1-0.05)=16.9190$，因此 H_0 不会被否决。因此我们可以认为 $\sigma^2=1.1$。实际上，只有样本均值 $s^2>2.0679$ 时 H_0 才会被否决。再考虑如下的假设：

$$H_0: \sigma^2 = \sigma_0^2, \quad H_1: \sigma^2 < \sigma_0^2 \tag{3-219}$$

对于这种情况，只有下面条件满足才能否决 H_0：

$$(N-1)S^2/\sigma_0^2 < F_Y^{-1}(\alpha_s) \tag{3-220}$$

因为式(3-218)的结果超过了 $F_Y^{-1}(0.05)=3.3251$，所以 H_0 不会被否决。实际上，需要样本均值 $s^2<0.4064$ 时 H_0 才会被否决。 ◀

3.8.18 *F* 分布

设$\{X_1, X_2\}$为独立的卡方分布随机变量，自由度分别为 m 和 n。那么下式服从 F 分布：

$$X = \frac{X_1/m}{X_2/n} \tag{3-221}$$

F 分布的概率密度函数为：

$$f_X(x)=\frac{\Gamma((m+n)/2)(m/n)^{m/2}x^{(m-2)/2}}{\Gamma(m/2)\Gamma(n/2)(1+(m/n)x)^{(m+n)/2}}I_{[0,\infty)}(x) \tag{3-222}$$

式(3-222)关于 m 和 n 并不是对称的。一个有趣的性质是，$1/X$ 同样也服从 F 分布，只不过自由度进行了相互交换，从上面的比值就可以看出。F 分布概率密度函数的例子如图 3-38所示。F 分布的累积分布函数为：

$$F_X(x)=B_{mx/(mx+n)}(m/2,n/2)I_{[0,\infty)}(x) \tag{3-223}$$

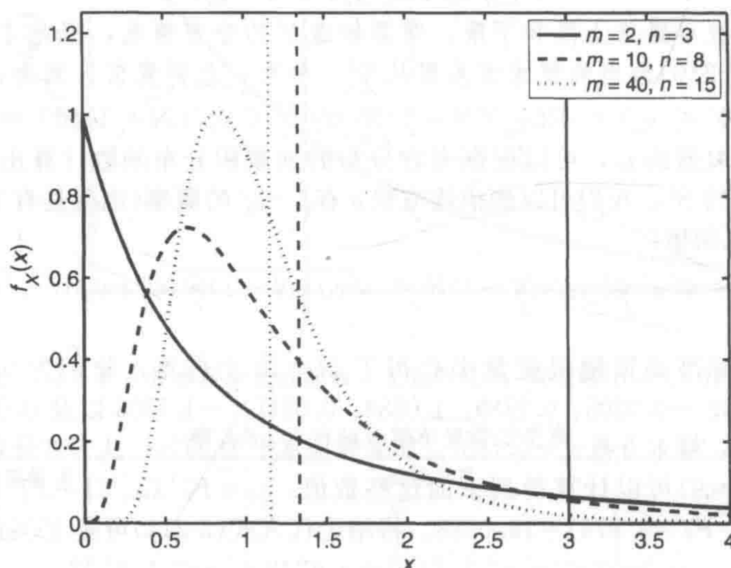


图 3-38 F 分布概率密度函数的例子，支集为$\mathcal{R}^+$，垂线指示了 μ_X 的位置

其中 $B_y(a,\ b)$为下不完全贝塔函数(见附录 B)。F 分布的均值为 $\mu_X=n/(n-2)$，$n>2$，方差为 $\sigma_X^2=2n^2(m+n-2)/[m(n-2)^2(n-4)]$，$n>4$。

例 3-50 与例 3-47 中的卡方分布随机变量类似，当 $m=\{1,2\}$ 时，F 分布的众数 m_o 也为 0，其中 m 为式(3-221)中分子里卡方分布随机变量的自由度。当 $m\geqslant 2$ 时，众数为(见习题 3-16)：

$$m_o=(m/n)(m-2)/(n+2) \tag{3-224}$$

F 分布经常通过下面的方法获取。假设从一个分布参数为$\{\mu_X,\ \sigma\}$的高斯分布中获取了 m 个样本$\{X_1,\cdots,X_m\}$，类似的，从另外一个高斯分布中获取 n 个样本$\{Z_1,\cdots,Z_n\}$，该高斯分布的方差仍然为 σ^2，均值为 μ_Z。设 $\overline{X}$ 和 $\overline{Z}$ 分别表示样本均值，并假设所有的样本都是相互独立的。那么下面这个随机变量就服从自由度为$\{m-1,n-1\}$的 F 分布：

$$Y\triangleq\left(\sum_{l=1}^{m}(X_l-\overline{X})^2/(m-1)\right)\left(\sum_{k=1}^{n}(Z_k-\overline{Z})^2/(n-1)\right)^{-1}=S_X^2/S_Z^2 \tag{3-225}$$

因为这个比值的每一项都是样本方差，因此可以将 Y 作为两个不同样本集合方差相似度的一种度量。◀

F 分布的应用场合包括：

- 利用假设检验，判断两个高斯随机变量各种的样本是否具有相同的方差。
- 评估最小均方(LS)模型的准确度，该模型用来给具有几何形式 $Y=aX+b$ 的样本建模，其中$\{a,b\}$是待估计的参数(最小均方估计将在第 9 章介绍)。

例 3-51 因为我们假设式(3-225)中基本分布的方差是相同的，可以用基于样本$\{X_n\}$和$\{Z_n\}$的随机变量 Y 来检验零假设 H_0：$\sigma_X^2=\sigma_Z^2$。如果另外一个假设是 H_1：$\sigma_X^2>\sigma_Z^2$，只有下面不等式成立时，H_0 才会被否决：

$$Y=S_X^2/S_Z^2>F_Y^{-1}(1-\alpha_s) \tag{3-226}$$

其中 $F_Y^{-1}(\cdot)$是 F 分布的逆累积分布函数，分布参数为$\{m-1,\ n-1\}$，α_s 是显著性水平。

设 $\sigma_X^2=2$，$\sigma_Z^2=1$，$m=n=10$。假设我们获得了 X 的以下样本：-0.3757，-2.2156，-0.5581，-0.2034，-3.3011，-1.9199，-2.5678，1.5675，-0.2010，1.5737，以及 Z 的样本：0.5593，0.4784，-0.6794，0.2850，-1.3329，-0.7240，-0.6636，0.1984，-1.7949，-1.3877。经计算样本均值 $\bar{x}=-0.8201$，$\bar{z}=-0.5061$，式(3-225)的比值为 $y=s_X^2/s_Z^2=2.7586/0.7167=3.8490$。因为 Y 超过了 $F_Y^{-1}(0.95)=3.1789$，因此 H_0 被否决。如果另外一个假设为 H_1：$\sigma_X^2<\sigma_Z^2$，那么当下面的不等式成立时，H_0 才会被否决：

$$1/Y = S_Z^2/S_X^2 > F_{1/Y}^{-1}(1-\alpha_s) \tag{3-227}$$

通过交换 $m-1$ 和 $n-1$，可以从 $F_Y^{-1}(\cdot)$ 得到逆累积分布函数 $F_{1/Y}^{-1}(\cdot)$。根据上面的数值结果，$1/y=0.2598$ 并没有超过 $F_{1/Y}^{-1}(0.95)=3.1789$，所以此时无法否决 H_0。请注意在这里 $F_Y^{-1}(\cdot)=F_{1/Y}^{-1}(\cdot)$，这是因为 $m=n$。◀

3.8.19 学生氏 t 分布

设 X_1 为高斯随机变量，X_2 为自由度为 $N=r$ 的卡方分布随机变量。对于独立的 X_1 和 X_2，下面的比值服从学生氏 t 分布：

$$X = X_1/\sqrt{X_2/r} \tag{3-228}$$

学生氏 t 分布的概率密度函数为：

$$f_X(x) = \frac{1}{\sqrt{r\pi}}\frac{\Gamma((r+1)/2)}{\Gamma(r/2)}\frac{1}{(1+x^2/r)^{(r+1)/2}}I_{(-\infty,\infty)}(x) \tag{3-229}$$

其中 $r\in\mathcal{N}$ 表示自由度。学生氏 t 分布存在两种特例：（ⅰ）当 $r=1$ 时，X 退化为分布参数为 $\{c=0,\ \alpha=1\}$ 的柯西随机变量，（ⅱ）当 $r\to\infty$ 时，得到标准高斯随机变量（见习题 4-2）。图 3-39给出了学生氏 t 分布概率密度函数的一些例子。它的累积分布函数为：

$$F_X(x) = B_{(x+\sqrt{x^2+r})/2\sqrt{x^2+r}}(r/2,r/2)I_{(-\infty,\infty)}(x) \tag{3-230}$$

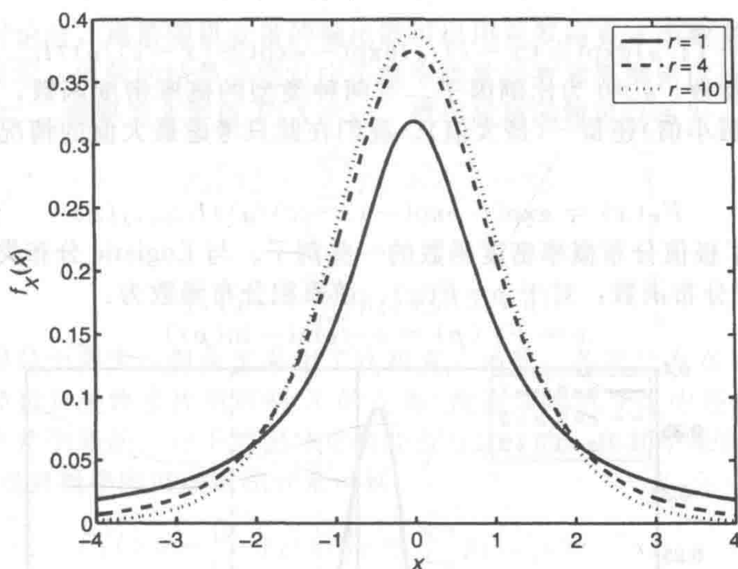


图 3-39 学生氏 t 分布概率密度函数的例子，支集为$\mathcal{R}$，当 $r>1$ 时 $\mu_X=0$，当 $r=1$ 时均值不存在

其中 $B_y(a,b)$ 为先前提到的下不完全贝塔函数。学生氏 t 分布的均值为 $\mu_X=0$，$r>1$，方差为 $\sigma_X^2=r/(r-2)$，$r>2$。

学生氏 t 分布可以通过下面的方法得到。设 $\{X_1,\cdots,X_r\}$ 是从分布参数为 $\{\mu,\sigma\}$ 的高斯分布中获取的样本。样本均值 $\overline{X}$ 同样也是高斯的，通过变换 $Y_1=\sqrt{r}(\overline{X}-\mu)/\sigma$ 可得标准高斯随机变量。根据前一节的讨论可知，$Y_2=(r-1)S^2/\sigma^2$ 服从自由度为 $r-1$ 的卡方分布。由于 Y_1 和 Y_2 是相互独立的，因此下面的比值服从学生氏 t 分布，分布参数为 $r-1$：

$$Y = \frac{Y_1}{\sqrt{Y_2/(r-1)}} = \frac{\overline{X}-\mu}{\sqrt{S^2/r}} \tag{3-231}$$

学生氏 t 分布的应用场合包括：

- 确定高斯随机变量样本均值的置信区间。
- 当存在较为严重的拖尾时，学生氏 t 分布可以替代高斯分布作为一个更加准确的模型(见表 3-4)。

例 3-52 与例 3-48 中使用的方法类似，基于卡方分布随机变量来推导方差的置信区间，我们可以将式(3-231)改写成如下形式：

$$P(y_l\sqrt{S^2/r} \leqslant \overline{X}-\mu \leqslant y_u\sqrt{S^2/r}) = P(\overline{X}-y_u\sqrt{S^2/r} \leqslant \mu \leqslant \overline{X}-y_l\sqrt{S^2/r}) \tag{3-232}$$

其中 $\{y_l, y_u\}$ 表示置信下限和上限，μ 前面的负号导致不等式改变了方向。将这个结果设为 $1-\alpha_c$，代入 $\overline{X}$ 和 S^2 可以得到 $100(1-\alpha_c)\%$ 置信区间，其中 $y_l=F_Y^{-1}(\alpha_c/2)$，$y_u=F_Y^{-1}(1-\alpha_c/2)$，$F_Y^{-1}(\cdot)$表示自由度为 $r-1$ 的学生氏逆累积分布函数。因为 $F_Y(y)$是关于原点对称的，可以代入 $y_l=-y_u$(对于例 3-48 中的置信限该结论不成立，因为卡方分布不是对称的)。假设我们在标准高斯分布中获得了如下 10 个样本：-0.8674，-0.4742，0.2224，1.8713，0.1100，-0.4113，0.5112，-1.1991，-0.0964 和 0.4458，可计算得到样本均值 $\overline{x}=0.0112$ 和样本方差 $s^2=0.7531$。对于 $\alpha_c=0.05$，有 $y_l=F_Y^{-1}(0.975)=-2.2622$ 以及 $y_u=F_Y^{-1}(0.025)=2.2622$，据此可以给出如下 95%的置信区间：

$$-0.6021 \leqslant \mu \leqslant 0.6246 \tag{3-233}$$

实际上，标准高斯分布 X 的均值为 $\mu=0$。◀

3.8.20　极值分布

有三种类型的极值分布，我们在此只讨论类型 I，也称作康拜尔分布。它的概率密度函数为：

$$f_X(x) = (1/\alpha)\exp(\pm(x-c)/\alpha)\exp(-\exp(\pm(x-c)/\alpha))I_{(-\infty,\infty)}(x) \tag{3-234}$$

其中 $c\in\mathcal{R}$为位置参数，$\alpha>0$ 为比例因子。有两种类型的概率密度函数，分别取决于指数前面的符号是+(最小值)还是−(最大值)。我们在此只考虑最大值的情况，此时累积分布函数为：

$$F_X(x) = \exp(-\exp(-(x-c)/\alpha))I_{(-\infty,\infty)}(x) \tag{3-235}$$

图 3-40 给出了极值分布概率密度函数的一些例子。与 Logistic 分布类似，可以直接推导得到它的逆累积分布函数：对于 $p=F(x)$，逆累积分布函数为：

$$x = F_X^{-1}(p) = c-\alpha\ln(-\ln(p)) \tag{3-236}$$

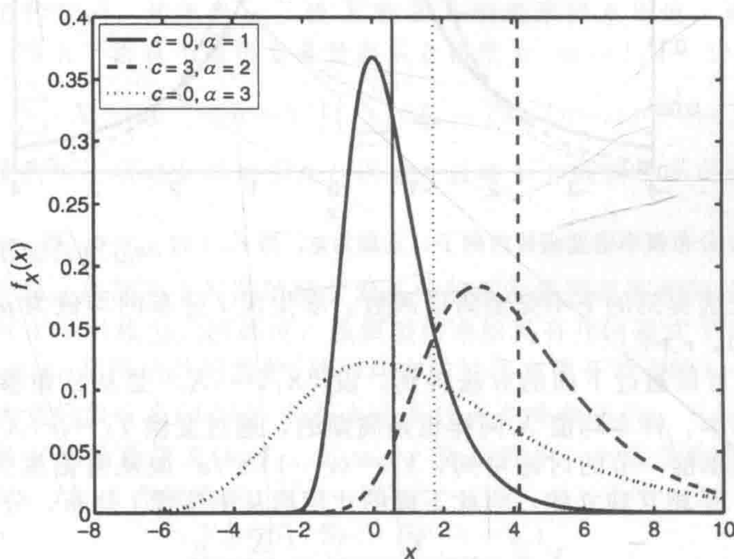


图 3-40　极值分布(康拜尔)概率密度函数的一些例子，支集为$\mathcal{R}$，垂线指示了 μ_X 的位置

极值分布的应用场合包括：

- 当 N 很大时，指数类型随机变量 N 个样本最大值的分布情况，也称作第 N 阶统计量(见第4章)。
- 对极端事件进行建模，比如最大地震运动、金属的断裂强度以及大风。

例 3-53 根据累积分布函数，可以用下面的方法计算中位数：

$$\begin{aligned}\exp(-\exp(-(m_e-c)/\alpha)) = 1/2 &\Rightarrow -\exp(-(m_e-c)/\alpha) = \ln(1/2) \\ &\Rightarrow -(m_e-c)/\alpha = \ln(\ln(2)) \\ &\Rightarrow m_e = c-\alpha\ln(\ln(2))\end{aligned} \tag{3-237}$$

因此，$m_e \approx c+(0.3665)\alpha$ 一般都会大于众数 $m_o=c$。均值为 $\mu_X=c+\alpha\gamma$，其中 γ 约为 0.5772，是欧拉-马歇罗马常数，可得 $\mu_X \approx c+(0.5772)\alpha$，均值一般大于中位数。从结果可知，概率密度函数偏向左边，而且斜率为正：$\gamma_1=(12\sqrt{6}/\pi^3)\zeta(3) \approx 1.1396$，其中 $\zeta(v)$ 为 Zeta 函数，$\zeta(3) \approx 1.2021$ 被称作 Apery 常数。康拜尔分布的方差为 $\sigma_X^2=(\pi\alpha)^2/6$。◀

例 3-54 可以通过变换 $X=c-\alpha\ln(Y)$ 得到康拜尔分布，其中 Y 是分布参数 $\lambda=1$ 的指数随机变量：

$$\begin{aligned}P(X \leqslant x) &= P([c-x]/\alpha \leqslant \ln(Y)) \\ &= 1-P(Y \leqslant \exp([c-x]/\alpha))\end{aligned} \tag{3-238}$$

代入指数随机变量的累积分布函数，可以得到我们需要的结果：

$$F_X(x) = 1-[1-\exp(-\exp(-[x-c]/\alpha))] \tag{3-239}$$

该结果即式(3-235)。◀

3.9 离散随机变量

在前面我们讨论过，离散随机变量的输出既可以用克罗内克 δ 函数来表示，也可以用狄拉克 δ 函数来表示，得到的结果分别对应于概率质量函数和概率密度函数。设离散随机变量 X 的输出为 $\mathcal{Z}^+$，概率为非零值 $P(X=x)$。概率质量函数可以写为：

$$p_X[x] = \sum_{n=0}^{\infty} p_X[n]\delta[x-n] \tag{3-240}$$

概率密度函数为：

$$f_X(x) = \sum_{n=0}^{\infty} p_X[n]\delta(x-n) \tag{3-241}$$

尽管求和上限是无限大，但由于采用了狄拉克 δ 函数，各项只有在 $n=x$ 时才是非零值。这里并不需要指示函数来特别声明 X 的支集(在连续随机变量中经常用到)，因为从求和表达式看是非常明确的。对于累积分布函数也是这样的，在其中采用的是单位阶跃函数 $u(x)$，可以通过对概率密度函数积分来得到：

$$F_X(x) = \int_{-\infty}^{x} f_X(v)\mathrm{d}v = \sum_{n=0}^{\infty} p_X[n]u(x-n) \tag{3-242}$$

为了对所有类型的随机变量有一致的描述方法(包括混合分布)，在本节中一般采用概率密度函数来进行表示。这样我们就可以通过积分来计算各种统计量，包括第5章将要介绍的均值：

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \sum_{n=0}^{\infty} p_X[n]\int_{-\infty}^{\infty} x\delta(x-n)\mathrm{d}x = \sum_{n=0}^{\infty} np_X[n] \tag{3-243}$$

在上面最后一个表达式中，用到了狄拉克 δ 函数的平移特性。在附录A中归纳了概率质量函数和概率密度函数的特性。下面对本节用到的各种符号进行一个归纳：

- p 表示“成功”的概率。
- $q \triangleq 1-p$ 表示“失败”的概率。

- N 表示试验/输出结果(二项、超几何、均匀)的数量，或者表示随机变量的求和(负二项)。
- M 表示成功对象的数目(超几何)。
- α 为一般性的比例因子(泊松和 zeta)。

3.9.1 伯努利随机变量

伯努利随机变量有两种输出结果，标记为 $\{0,1\}$，概率密度函数为：

$$f_X(x) = q\delta(x) + p\delta(x-1) \tag{3-244}$$

其中 $p = P(X=1) \in [0,1]$ 表示成功(比如在硬币投掷试验中正面朝上)的概率，$q \triangleq 1-p$ 表示失败的概率。对称伯努利随机变量的输出为 $\{-1,1\}$（见附录 A)。用单位阶跃函数替代狄拉克 δ 函数就可得到伯努利累积分布函数：

$$F_X(x) = qu(x) + pu(x-1) \tag{3-245}$$

图 3-41 给出了伯努利分布概率密度函数的一些例子。均值为 $\mu_X = p$，方差为 $\sigma_X^2 = pq$。伯努利分布的应用场合包括：

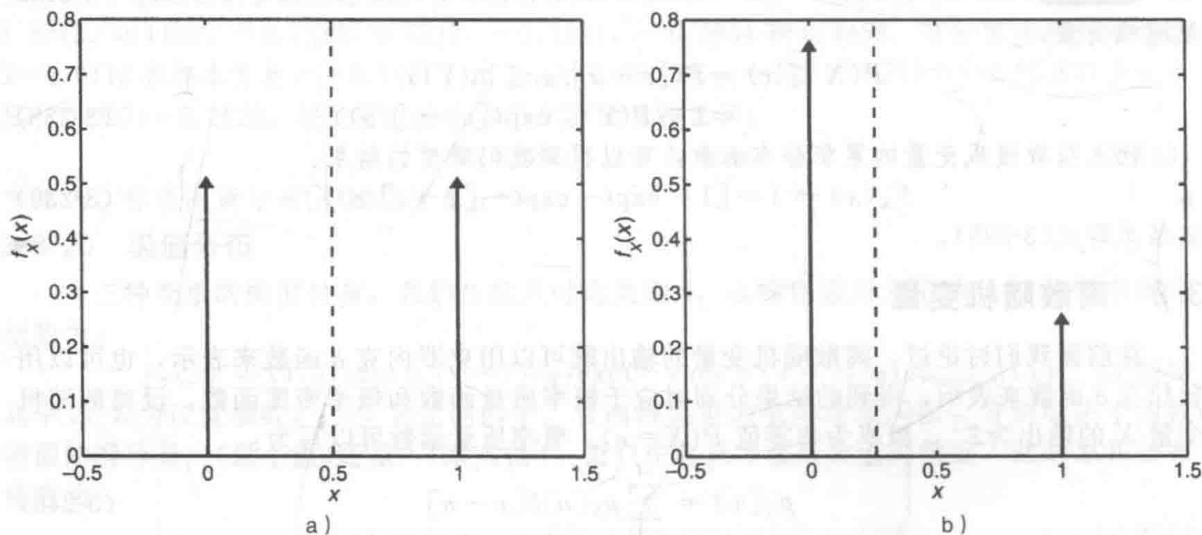


图 3-41 伯努利概率密度函数的一些例子。a) $p=0.5$；b) $p=0.25$。支集为 $\{0,1\}$，虚线指示了 μ_X 的位置

- 输出结果为 $\{H,T\}$ 形式的许多试验，比如单个硬币投掷试验，或者互补事件比如投掷骰子，输出结果为偶数或奇数。
- 数字通信系统中的二进制数字(bit)。

例 3-55 对于绝对概率空间 $\{\Omega,\mathcal{F},P\}$，其中 Ω 可能是不可数的。设随机指示变量可以将某个事件 $E\in\mathcal{F}$映射到$\mathcal{R}$：

$$X(\omega) = I_E(\omega) \tag{3-246}$$

X 只有两种输出，因此 X 是一个贝努利随机变量：

$$p = P(X=1) = P(E), q = P(X=0) = P(E^c) \tag{3-247}$$

假设 Ω 表示单次投掷骰子试验的输出。例如，如果 X 表示输出结果大于 4 的事件，那么 X 仍然是一个伯努利随机变量，且 $p=1/3$，$q=2/3$。 ◀

例 3-56 (熵)离散随机变量的熵是对输出分布中携带的信息或者不确定性的一种度量。它的定义为：

$$H(X) = -\sum_{n=0}^{N} p_X[n]\log(p_X[n]) \tag{3-248}$$

对数函数可以取任何数为底，但我们一般都以 2 或 e 为底，此时 $H(X)$ 的单位为比特(bit)

或莱特(nat)。在第 10 章还将对此进行详细介绍。在此，我们只简单的考虑一下伯努利随机变量的熵：

$$H(X) = p\log(p) + q\log(q) \tag{3-249}$$

图 3-42 给出了式(3-249)的图形。可以看出当 $p=1/2$ 时(最大的不确定性)$H(X)$取最大值，当 $p=\{0,1\}$ 时(没有任何不确定性)$H(X)$取最小值。后一种情况的信息量为 0，因为此时我们能够以 1 的概率分别知道输出为 0 或 1。

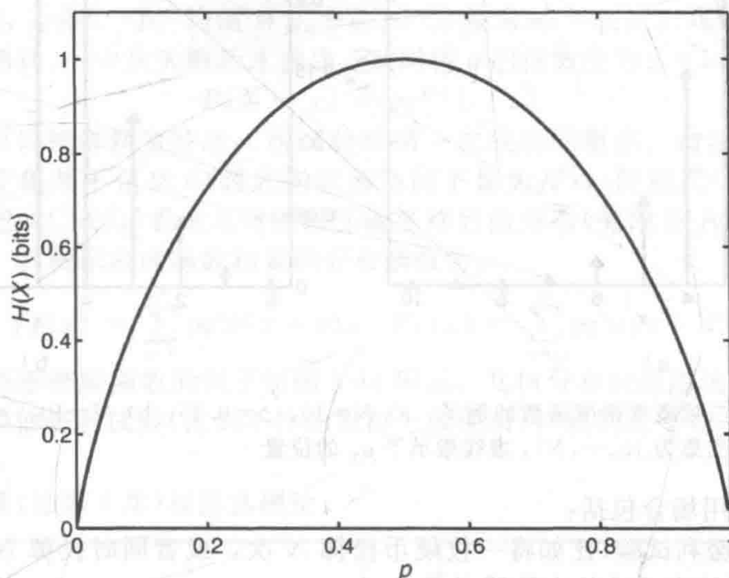


图 3-42　伯努利随机变量的熵(比特)

3.9.2　二项随机变量

将伯努利试验重复 N 次(全部都是独立的)后统计成功的次数，得到的就是二项随机变量。二项分布概率密度函数为：

$$f_X(x) = \sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}\delta(x-n) \tag{3-250}$$

其中 $p\in[0,\ 1]$，$q\triangleq 1-p$。下式为二项式系数，它表示 N 次伯努利试验中成功 n 次的所有可能的排列数目：

$$\binom{N}{n}\triangleq\frac{N!}{n!(N-n)!} \tag{3-251}$$

上面关于二项分布的描述对应于伯努利随机变量的替换性采样，已经在第 2 章中关于有限样本空间 Ω 进行了讨论。样本规模为 N 的二项概率可由下面的递归表达式进行计算(见习题 3-37)：

$$P(X=x) = [(N-x+1)p/qx]P(X=x-1) \tag{3-252}$$

累积分布函数为：

$$F_X(x) = \sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}u(x-n) \tag{3-253}$$

对于一个特定取值的 $x<N$，只有 $n\leqslant x$ 时单位阶跃函数才是非零的。求和表达式中$n>x$ 的部分都被排除在外(等于 0)。我们也可以不用单位阶跃函数来写出累积分布函数的表达式：

$$F_X(x) = \sum_{n=0}^{\lfloor x\rfloor}\binom{N}{n}p^n q^{N-n} \tag{3-254}$$

其中$\lfloor x\rfloor$为下取整函数，该函数返回小于 x 的最大整数。利用这个记号可以很好的定义求和

表达式，对于非整数取值的x累积分布函数为一个常数，这正是离散随机变量需要的。二项随机变量的概率密度函数如图 3-43 所示。均值 $\mu_X = Np$，方差 $\sigma_X^2 = Npq$。

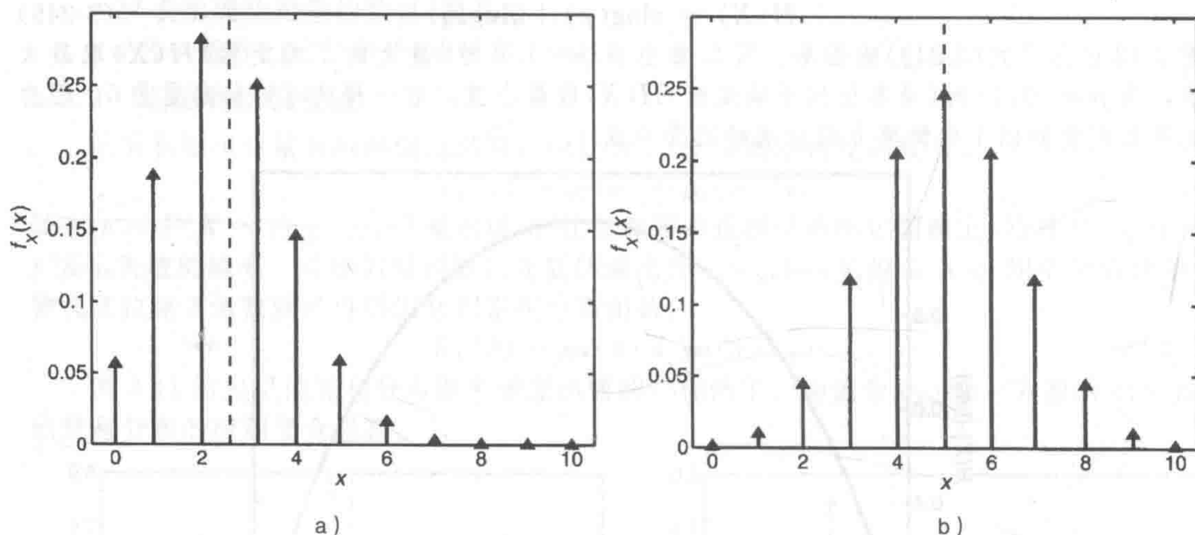


图 3-43 二项概率密度函数的例子。a) $N=10$，$p=0.25$；b) $N=10$，$p=0.5$。支集为 $\{0,\cdots,N\}$，虚线指示了 μ_X 的位置

二项分布的应用场合包括：

- 重复进行伯努利试验(比如将一枚硬币投掷 N 次，或者同时投掷 N 枚硬币)，我们对这 N 次试验中成功的次数感兴趣。
- 对于较大的样本规模 N，预测次品的数量。

例 3-57 将一枚硬币投掷 $N=3$ 次，概率密度函数为：

$$f_X(x) = \sum_{n=0}^{3}\binom{3}{n}(1/2)^n(1/2)^{3-n}\delta(x-n) = (1/8)\sum_{n=0}^{3}\binom{3}{n}\delta(x-n) \tag{3-255}$$

假设我们希望计算 $P(X=3)$。因为式(3-250)是一个概率密度函数，概率可以用 $x=3$ 为中心的一个小区间上的积分来进行计算：

$$P(X=3) = \int_{3-\varepsilon}^{3+\varepsilon} f_X(x)\mathrm{d}x = (1/8)\binom{3}{3} = 1/8 \tag{3-256}$$

其中很小的 $\varepsilon \in (0, 1)$可以帮助我们“捕捉”到 $x=3$ 处的 δ 函数。◀

现考虑另外一个例题：

$$P(X>1) = \int_{1+\varepsilon}^{\infty} f_X(x)\mathrm{d}x = (1/8)\left[\binom{3}{2}+\binom{3}{3}\right] = (1/8)(3+1) = 1/2 \tag{3-257}$$

式(3-257)选择了合适的积分下限，使得输出结果$\{0, 1\}$被排除在外。同样也可以根据累积分布函数进行计算：

$$P(X>1) = 1-P(X\leqslant 1) = 1-(1/8)\sum_{n=0}^{3}\binom{3}{n}u(1-n) \tag{3-258}$$

上述求和表达式中有 4 项，其中 $u(0)=u(1)=1$，而 $u(-2)=u(-1)=0$，因此可得：

$$P(X>1) = 1-(1/8)\left[\binom{3}{0}+\binom{3}{1}\right] = 1/2 \tag{3-259}$$

例 3-57 介绍了如何根据概率密度函数和累积分布函数计算离散随机变量的概率，需要分别基于狄拉克 δ 函数和单位阶跃函数来进行计算。但是，读者可能会发现用概率质量函数来计算在概念上会更简单一些，只需要将各个概率简单相加即可。对概率质量函数进行求和与对概率密度函数进行积分，二者其实是一致的，因为概率密度函数中的狄拉克 δ 函数与概率质量函数中的克罗内克 δ 函数都位于 X 的相同输出结果处。

3.9.3 几何随机变量(支集为$\mathcal{Z}^+$或$\mathcal{N}$)

几何随机变量常用来定义伯努利试验中经过 x 次失败后第一次成功需要的试验次数。有的时候它也被称作"等待时间"随机变量，在这里用失败的次数来定义"时间"(并不是真正意义上的时间)。因为每次试验都是独立的，我们可以写成下面的表达式：

$$P(X=x)=\underbrace{q\cdots q}_{x\text{ 项}}p=pq^{x}I_{\mathcal{Z}^+}[x] \tag{3-260}$$

式中，$p\in[0,1]$，$q\triangleq 1-p$。均值为 $\mu_X=q/p$，方差为 $\sigma_X^2=q/p^2$。几何随机变量的另外一种定义为：观测到 $x-1$ 次失败后才成功，此时将 q 的指数改为 $x-1$，可得：

$$P(X=x)=pq^{x-1}I_{\mathcal{N}}[x] \tag{3-261}$$

几何分布还可以被解释为经过 x 次试验后第一次成功的概率。请注意：式(3-261)定义的几何分布的支集并不包括 0(因此指示函数的下标为$\mathcal{N}$)，但在式(3-260)中是包括 0 的。因此，我们把式(3-261)看做几何随机变量平移后的分布(见附录 A)。式(3-260)定义的几何随机变量，其概率密度函数和累积分布函数为：

$$f_X(x)=\sum_{n=0}^{\infty}pq^{n}\delta(x-n),\quad F_X(x)=\sum_{n=0}^{\infty}pq^{n}u(x-n) \tag{3-262}$$

几何分布的概率密度函数的例子如图 3-44 所示。几何分布的应用场合包括：

- 重复进行的伯努利试验(比如不停的投掷一枚硬币)，我们对 x 次失败后的第一次成功感兴趣。
- 马尔可夫链(见第 6 章)和排队理论。

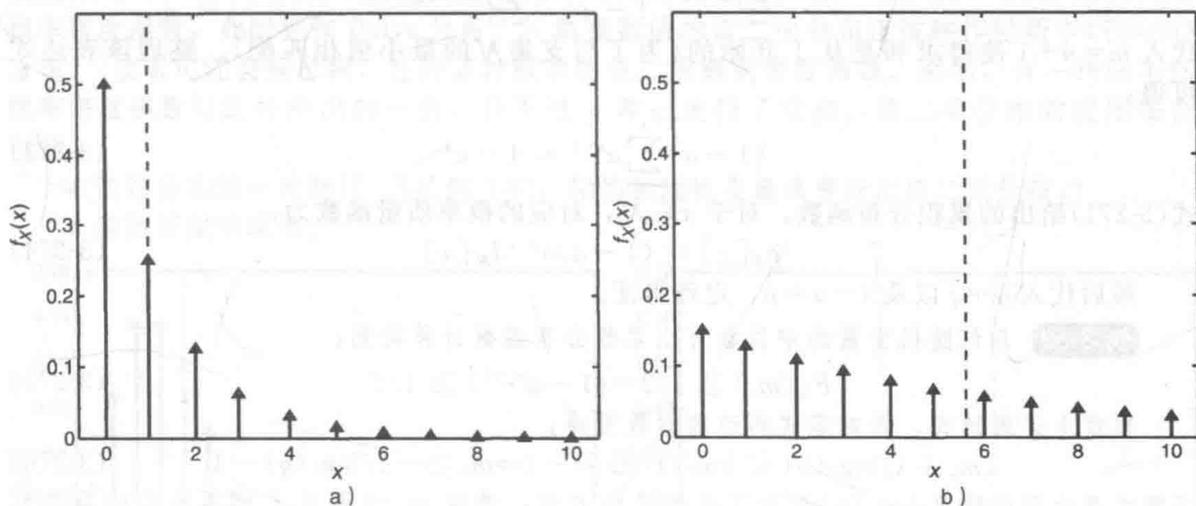


图 3-44 几何分布概率密度函数的例子。a) $p=0.5$；b) $p=0.15$。支集为$\mathcal{Z}^+$，虚线指示了 μ_X 的位置

例 3-58 对于一个比特字符串，输出为 $\{0,1\}$ 且都是等概率的。当 $n\in\{0,\cdots,x\}$ 时 $u(x-n)=1$，当 $n>x$ 时 $u(x-n)=0$，因此有：

$$P(X>x)=1-F_X(x)=1-p\sum_{n=0}^{\infty}q^{n}u(x-n)=1-p\sum_{n=0}^{x}q^{n} \tag{3-263}$$

正如在例 3-57 中曾提到，将概率质量函数中的对应项进行累加就可得到计算结果。根据序列求和结果的闭合表达式可得(见附录 E)：

$$P(X>x)=1-p(1-q^{x+1})/(1-q)=q^{x+1} \tag{3-264}$$

与其他类型的离散随机变量不同，其累积分布函数有如下简单的闭合表达式：

$$F_X(x)=1-P(X>x)=(1-q^{x+1})I_{\mathcal{Z}^+}[x] \tag{3-265}$$

可以看出，累积分布函数在 x 的两个整数取值之间是平坦的。 ◀

定理 3-4 集合随机变量是唯一具有无记忆特性的离散随机变量。

证明：根据条件概率可知：

$$P(X>x+b|X>b)=P(X>x+b)/P(X>b)=[1-F_X(x+b)]/[1-F_X(b)] \tag{3-266}$$

代入几何随机变量的累积分布函数可得：

$$P(X>x+b|X>b)=q^{x+b+1}/q^{b+1}=q^x\neq 1-F_X(x) \tag{3-267}$$

可以看出，式(3-260)定义的概率密度函数并不具备无记忆特性(式 3-87)。现在再考察式(3-261)给出的平移几何随机变量，其累积分布函数为：

$$F_X(x)=(1-q^x)I_{\mathcal{N}}[x] \tag{3-268}$$

将上述定义代入条件概率表达式，可得：

$$P(X>x+b|X>b)=q^{x+b}/q^b=q^x=1-F_X(x) \tag{3-269}$$

可以看出，支集为$\mathcal{N}$的平移几何随机变量的概率密度函数具有无记忆特性(0 必须被排除在外)。

很显然，对于整数 x，只有 a^x 的函数形式才能满足 $a^{x+b}/a^b=a^x$，因此这也是唯一满足下面不等式的函数表达式：

$$P(X>x+b)/P(X>b)=P(X>x) \tag{3-270}$$

因此可得：

$$P(X>x)=a^x\Rightarrow P(X\leqslant x)=1-a^x=F_X(x) \tag{3-271}$$

从式(3-271)可以看出，在 $x=0$ 处的概率质量为 0，$a\in[0,1]$ 是为了确保 $1-a^x$ 是一个合法的分布。对于有限项求和，考虑如下的闭合表达式(见附录 E)：

$$\sum_{n=0}^{x-1}a^n=\frac{1-a^x}{1-a}=\sum_{m=1}^{x}a^{m-1} \tag{3-272}$$

代入 $m=n+1$ 使得求和是从 1 开始的(为了与支集$\mathcal{N}$的最小值相匹配)。整理该表达式可得：

$$(1-a)\sum_{m=1}^{x}a^{m-1}=1-a^x \tag{3-273}$$

式(3-271)给出的累积分布函数。对于 $x\in\mathcal{N}$，对应的概率质量函数为：

$$p_X[x]=(1-a)a^{x-1}I_{\mathcal{N}}[x] \tag{3-274}$$

最后代入 $a=q$ 以及 $1-a=p$，定理得证。

例 3-59 几何随机变量的中位数可由累积分布函数计算得到：

$$F_X(m_e)\geqslant 1/2\Rightarrow(1-q^{m_e+1})\geqslant 1/2 \tag{3-275}$$

整理上述表达式，并对等式两边取对数可得：

$$(m_e+1)\log_2(q)\leqslant\log_2(1/2)=-1\Rightarrow m_e\geqslant -1/\log_2(q)-1 \tag{3-276}$$

不等式两边同时除以 $\log_2(q)$ 改变了表达式的方向。选取 m_e 为满足上面不等式的最小整数，m_e 的计算可用如下的上取整函数(ceiling function，见附录 B)：

$$m_e=\lceil -1/\log_2(q)\rceil-1 \tag{3-277}$$ ◀

表 3-6 对二项随机变量、几何随机变量与伯努利随机变量的关系进行了归纳。表 3-6 还包括负二项随机变量，它是几何分布在 $N=1$ 时的特例。

表 3-6 伯努利试验和相关的随机变量

随机变量 X	分布
N 次试验中 x 次成功	二项分布 $\{N,p\}$
x 次失败后的第一次成功	几何分布$\{p\}$(支集为$\mathcal{Z}^+$)
x 次试验后的第一次成功	平移几何分布$\{p\}$(支集为$\mathcal{N}$)
x 次失败后的第 N 次成功	负二项分布 $\{N,p\}$

3.9.4 负二项随机变量(帕斯卡)

负二项随机变量帕斯卡(Pascal)的产生过程为：将 N 个几何随机变量进行累加，每个

变量都服从相同的分布参数 $p\in[0,1]$，计算经过 x 次失败后出现 N 次成功的概率。伯努利输出结果的一个可能序列为：第 1 次失败，然后随后的 N 次全是成功，这种情况的概率为 $p^N q$，$q\triangleq 1-p$。在前 $N+1$ 次试验中发生 $x=1$ 次失败的排列数目为 $\binom{N}{1}=N$：

$$\underbrace{\overbrace{\{0,1,\cdots,1\}}^{N+1\text{项}},\{1,0,1,\cdots,1\},\cdots,\{1,\cdots,1,0,1,1\},\{1,\cdots,1,0,1\}}_{N\text{种排列}} \tag{3-278}$$

将这个结果应用于 x 次失败的情况，对于所有的排列：

$$P(X=x)=\binom{x+N-1}{x}p^N q^x I_{\mathcal{Z}^+}[x] \tag{3-279}$$

请注意，对于 $x=0$ 次失败的情况只存在一种排列。根据定义，成功的数量一定是 N，N 次成功之前需要 x 次失败，排列数量可由二项式系数给出(很显然，无论 N 取多少，最后一次的输出结果一定是成功)。因此，负二项随机变量的概率密度函数为：

$$f_X(x)=\sum_{n=0}^{\infty}\binom{n+N-1}{n}p^N q^n\delta(x-n) \tag{3-280}$$

累积分布函数为：

$$F_X(x)=\sum_{n=0}^{\infty}\binom{n+N-1}{n}p^N q^n u(x-n) \tag{3-281}$$

负二项概率密度函数的例子如图 3-45 所示。均值为 $\mu_X=Nq/p$，方差 $\sigma_X^2=Nq/p^2$。很显然，当 $N=1$ 时，负二项随机变量退化为几何随机变量。也可以生成 $N\in\mathcal{R}^+$ 的负二项概率密度函数，此时称作 Polya 分布。N 取整数值的负二项分布也被称作帕斯卡(Pascal)分布。(读者可能会注意到，有许多种版本的负二项概率密度函数。例如，有一种版本的概率密度函数与此处给出的一致，只不过 p 和 q 进行了互换。)负二项分布的应用场合包括：

- 泊松分布的一种替代。(见例 3-62，用泊松随机变量来推导出负二项分布。)
- 事故征候的模型。

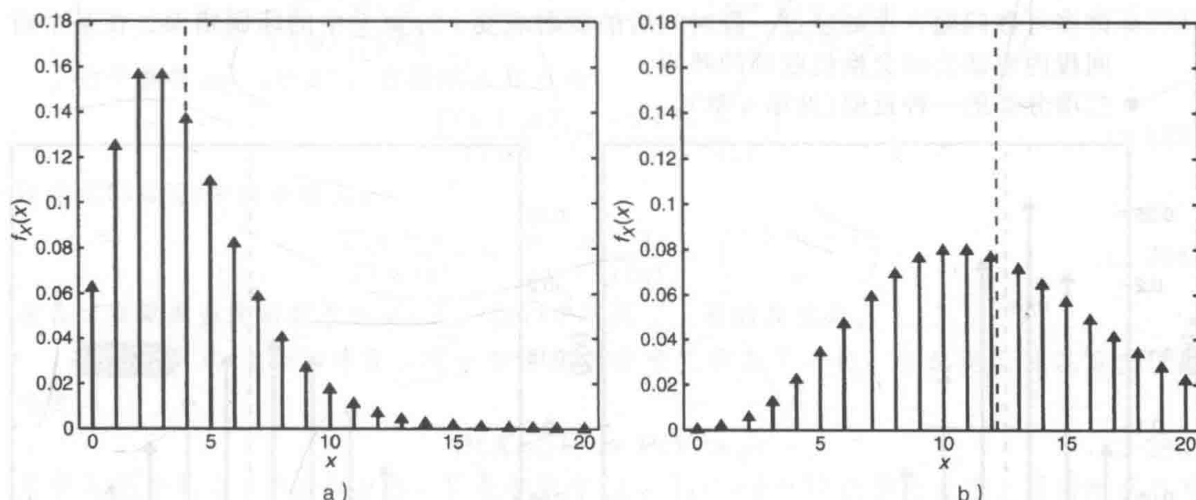


图 3-45 负二项概率密度函数的例子。a) $N=4$，$p=0.5$；b) $N=10$，$p=0.45$。支集为$\mathcal{Z}^+$，虚线指示了 μ_X 的位置

例 3-60 设 $N=2$，$p=q=1/2$。对于非常小的 $\varepsilon\in(0,1)$，有下面的结论：

$$\begin{aligned}P(2\leqslant X\leqslant 4)&=\int_{2-\varepsilon}^{4+\varepsilon}f_X(x)\mathrm{d}x=p^2\left[\binom{3}{2}q^2+\binom{4}{3}q^3+\binom{5}{4}q^4\right]\\&=p^2[2q^2+4q^3+5q^4]=(1/2)[1/2+3/8+5/16]\end{aligned} \tag{3-282}$$

以硬币投掷试验为例，得到在 $N=2$ 次正面前出现 $x=2$，3，4 次背面的概率为 $19/32\approx 0.5938$。◀

例 3-61 (Krishnamoorthy，2006)设 X 服从分布参数为$\{N, p\}$的负二项分布。那么对于整数 k：

$$P(X \leqslant k) = P(Y \geqslant N) \tag{3-283}$$

其中 Y 是分布参数为 $\{N+k, p\}$ 的负二项随机变量。为了将概率表达式写在等式左侧，我们认为事件 E 和 F 是等价的，其中事件 E 表示"第 N 次成功前少于或等于 k 次失败"，事件 F 表示"获得 N 次成功，试验次数少于或等于 $k+N$"。对于二项随机变量，$k+N$ 次试验中有 n 次成功的概率为：

$$P(Y = n) = \binom{k+N}{n} p^n q^{k+N-n} \tag{3-284}$$

事件 G 表示"进行了 $k+N$ 次试验，成功次数大于或等于 N"。尽管 F 和 G 是不同的事件，但关键的一步就是证明这两个事件的概率相同。因此有：

$$P(G) = \sum_{n=N}^{k+N} \binom{k+N}{n} p^n q^{k+N-n} \tag{3-285}$$

可知这个概率与 $P(Y \geqslant N)$ 是相同的。◀

3.9.5 泊松随机变量

在许多场合都会遇到泊松随机变量，其概率密度函数为：

$$f_X(x) = \sum_{n=0}^{\infty} \exp(-\alpha) \frac{\alpha^n}{n!} \delta(x-n) \tag{3-286}$$

其中 $\alpha > 0$ 是位置参数，同时也是形状参数。累积分布函数为：

$$F_X(x) = \sum_{n=0}^{\infty} \exp(-\alpha) \frac{\alpha^n}{n!} u(x-n) \tag{3-287}$$

泊松分布概率密度函数的一些例子如图 3-46 所示。均值和方差相等：$\mu_X = \sigma_X^2 = \alpha$。泊松分布的应用场合包括：

- 许多可数问题，比如经过一段时间后的放射衰变、一页书中的印刷错误、在某个时间段内电话公司交换机收到的呼叫。
- 二项分布的一种近似(见第 4 章)。

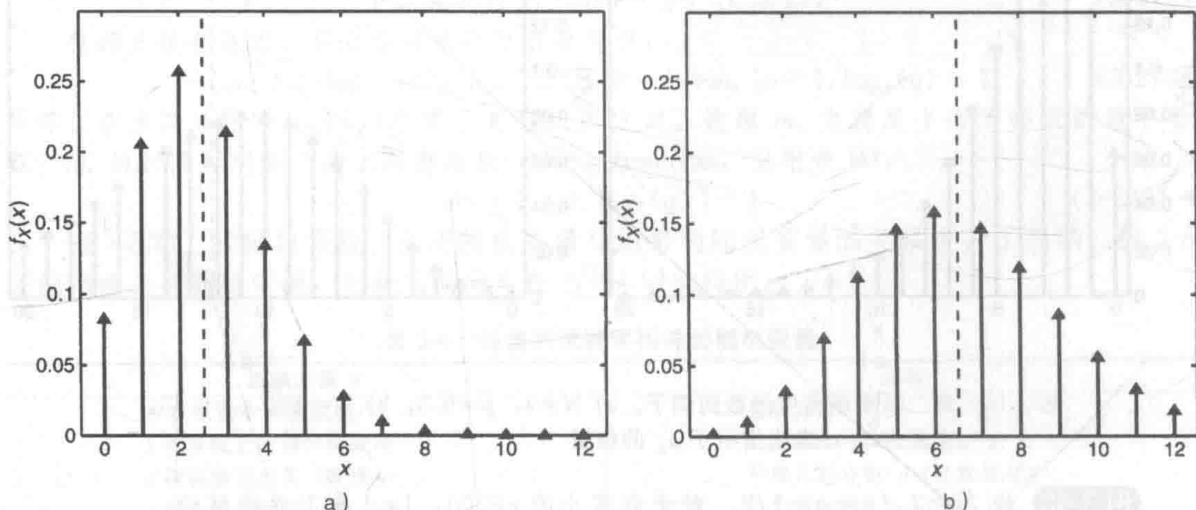


图 3-46 泊松概率密度函数的一些例子。a) $\alpha=2.5$；b) $\alpha=6.5$。支集为$\mathcal{Z}^+$，虚线指示了 μ_X 的位置

我们在第 6 章要讨论泊松计数过程，泊松概率密度函数的表达式包含了该过程汇总时

间t，因此α变成$\alpha=\lambda t$，λ表示速率参数。

例 3-62 当α服从参数为q/p的伽马分布时，泊松分布可以被用来推导负二项分布。因为α此时是随机的，可以用条件概率来写出泊松概率密度函数表达式（见第4章关于条件概率的讨论）：

$$f_{X|\alpha}(x|\alpha)=\frac{\alpha^x}{x!}\exp(-\alpha)I_{[0,\infty)}(x) \tag{3-288}$$

（因为α的大写与拉丁字母A相同，我们在此采用小写的α，这样不会引起混淆，它仍然表示一个随机变量。）对α取条件概率意味着α取了一个特定值，此时α不再是随机的。将α的概率密度函数代入下面积分表达式，并在它的支集上进行积分，得到X的概率密度函数为：

$$\int_0^\infty f_{X|\alpha}(x|\alpha)f_\alpha(\alpha)\mathrm{d}\alpha=\int_0^\infty f_{X,\alpha}(x,\alpha)\mathrm{d}\alpha=f_X(x) \tag{3-289}$$

式(3-289)为一个混合分布，$\{f_{X|\alpha}(x|\alpha)\}$的分布参数为$\alpha$（在第4章还会讲到混合分布）。$f_{X|\alpha}(x|\alpha)$是参数为$\alpha$的泊松分布，$f_\alpha(\alpha)$是参数为$\{r=m,\ \lambda=p/q\}$的伽马概率密度函数，可知$f_X(x)$是参数为$\{p,\ N=m\}$的负二项概率密度函数：

$$\begin{aligned}f_X(x)&=\int_0^\infty\sum_{n=0}^\infty\frac{\alpha^n}{n!}\exp(-\alpha)\delta(x-n)\frac{1}{\Gamma(m)}\alpha^{m-1}(p/q)^m\exp(-\alpha p/q)\mathrm{d}\alpha\\&=\frac{(p/q)^m}{\Gamma(m)}\sum_{n=0}^\infty\frac{1}{n!}\delta(x-n)\int_0^\infty\alpha^{n+m-1}\exp(-\alpha/q)\mathrm{d}\alpha\end{aligned} \tag{3-290}$$

在指数运算中用到了$1+p/q=1/q$。如果代入$v=\alpha/q$和$\mathrm{d}\alpha=q\mathrm{d}v$，积分表达式成为一个伽马函数：

$$\begin{aligned}\int_0^\infty\alpha^{n+m-1}\exp(-\alpha/q)\mathrm{d}\alpha&=q^{n+m}\int_0^\infty v^{n+m-1}\exp(-v)\mathrm{d}v\\&=q^{n+m}\Gamma(n+m)\end{aligned} \tag{3-291}$$

因此可得：

$$f_X(x)=\frac{p^m}{\Gamma(m)}\sum_{n=0}^\infty\frac{q^n}{n!}\Gamma(n+m)\delta(x-n)=\sum_{n=0}^\infty\frac{\Gamma(n+m)}{\Gamma(m)n!}p^mq^n\delta(x-n) \tag{3-292}$$

对于整数m，$x\in\mathcal{Z}^+$，根据附录B可知：

$$\frac{\Gamma(n+m)}{\Gamma(m)}=\frac{(n+m-1)!}{(m-1)!} \tag{3-293}$$

因此式(3-292)中的系数为：

$$\frac{\Gamma(n+m)}{\Gamma(m)n!}=\frac{n+m-1}{(m-1)!n!}=\binom{n+m-1}{n} \tag{3-294}$$

在负二项概率密度函数表达式（式3-280）中用到了上面的表达式。◀

例 3-63 式(3-283)将负二项分布与二项分布联系在了一起，与此类似可以写出下面等式：

$$P(X\leqslant k)=P(Y>\alpha) \tag{3-295}$$

其中X是参数为α的泊松分布，Y是参数为$\{\lambda=1,r=k+1\}$的伽马分布。根据附录B可知，下不完全伽马函数可以写成下面的级数形式：

$$\gamma(a,b)=\exp(-b)\sum_{n=0}^\infty\frac{(a-1)!x^{a+b}}{(a+n)!} \tag{3-296}$$

式(3-296)为伽马随机变量Y累积分布函数的分布参数，因此可得：

$$P(Y>\alpha)=1-F_Y(\alpha)=1-\frac{\gamma(k+1,\alpha)}{\Gamma(k+1)} \tag{3-297}$$

代入$\Gamma(k+1)=k!$和式(3-296)可得：

$$P(Y>\alpha)=1-\exp(-\alpha)\frac{1}{k!}\sum_{n=0}^{\infty}\frac{k!\alpha^{k+1+n}}{(k+1+n)!} \tag{3-298}$$

分子分母同时消掉$k!$，将求和表达式中的变量替换为$m=k+1+n$后得到最终结果：

$$P(Y>\alpha)=1-\exp(-\alpha)\sum_{m=k+1}^{\infty}\frac{\alpha^m}{m!}=\exp(-\alpha)\sum_{m=0}^{k}\frac{\alpha^m}{m!}=p(X\leqslant k) \tag{3-299}$$ ◀

3.9.6 超几何随机变量

超几何随机变量与二项随机变量比较类似，但有两个区别：(ⅰ)有$N\in\mathcal{N}$个对象可供选择。(ⅱ)采样是非替代性的。也就是说某个对象被选取以后，该对象就被移除在后续的选取范围之外了。对于所有的待选对象存在两种类型：一种的规模是M(成功)，另外一种的规模是$N-M$(失败)。并不需要非得把这些对象命名为"成功"或"失败"，实际只需要对它们二者进行区分即可，对其中一种感兴趣的对象命名为"成功"。如果有n个对象被选出，不考虑对象类型的话存在$\binom{N}{n}$种组合。我们在此研究m个对象中出现n次成功的数目，对于下面的这个场景：

$$\underbrace{1,\cdots,1}_{M项}\ \underbrace{0,\cdots,0}_{N-M项} \tag{3-300}$$

其中1表示成功，0表示失败。因为存在M个成功对象，因此有$\binom{M}{m}$种组合。同样可知，有$\binom{N-M}{n-m}$种失败对象的组合。因此，m个成功的概率为下面的比值：

$$P(X=m)=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \tag{3-301}$$

上面的结果可以根据组合数学计算：我们先统计实现m次成功和$n-m$次失败的所有情况的数量，然后再除以可以选取的n个对象的数量。(这个方法在第2章中也用到过，输出结果为等概率的Ω，用某个事件E的基数来计算比值$P(E)=|E|/|\Omega|$。)超几何随机变量的概率密度函数为：

$$f_X(x)=\sum_{m=\max(0,n+M-N)}^{\min(M,n)}\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}\delta(x-m) \tag{3-302}$$

超几何概率密度函数的例子如图3-47所示。求和表达式中的限制是有必要的，这样确保每个二项式系数的求和下限不会超过求和上限。如果二项式系数可以包括0，那么求和上下限就可以简化为$\{0,N\}$。累积分布函数为：

$$F_X(x)=\sum_{m=\max(0,n+M-N)}^{\min(M,n)}\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}u(x-m) \tag{3-303}$$

超几何随机变量X的均值为$\mu_X=nM/N$，方差$\sigma_X^2=nM(N-M)(N-n)/N^2(N-1)$。超几何分布的应用场合包括：

- 产品的质量控制，即通过抽样来判断次品率。
- 估计一种动物的数量，即对M个的样本"打标签"，然后将这部分样本放回种群，再次进行采样(n个样本)来估计总数M。

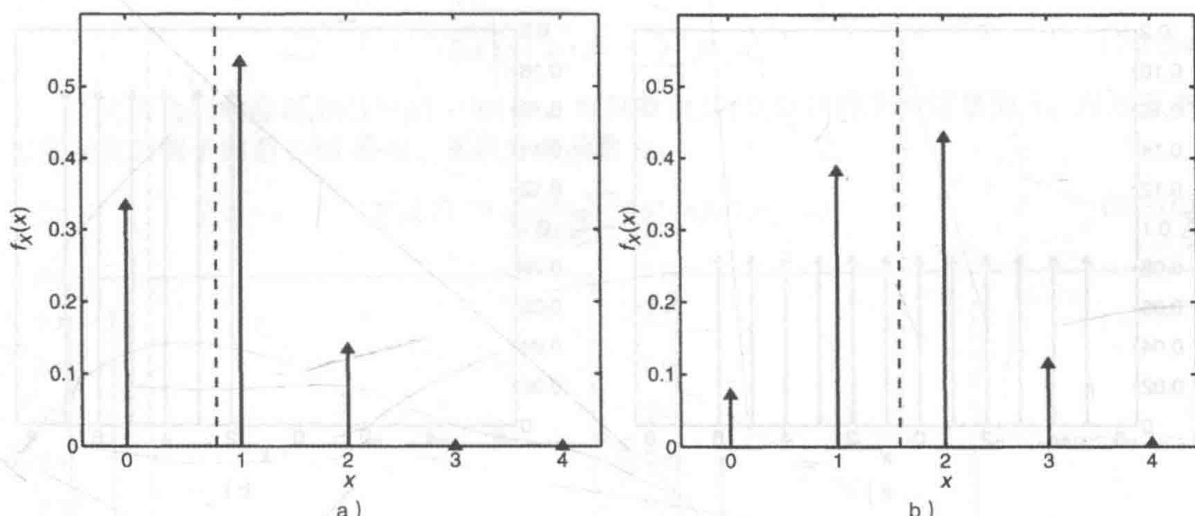


图 3-47 超几何概率密度函数的例子。a) $N=10$，$M=4$，$n=2$；b) $N=10$，$M=n=4$。支集为 $\{\max(0,n+M-N),\cdots,\min(M,n)\}$，虚线指示了 μ_X 的位置

例 3-64 研究 $n=1$ 的情况，此时概率密度函数简化为：

$$f_X(x)=\sum_{m=0}^{1}\frac{\binom{M}{m}\binom{N-M}{1-m}}{\binom{N}{1}}\delta(x-m)$$
$$=[(N-M)/N]\delta(x)+(M/N)\delta(x-1) \tag{3-304}$$

如果我们定义 $p\triangleq M/N$，那么对于 $n=1$ 的情况超几何分布就退化为伯努利分布。在第 4 章会讲到，当 M 和 N 取值都很大时，无论 n 的取值是多少，用这种方法定义的二项分布(参数为 p)可以近似为超几何分布。◀

例 3-65 设一艘船上的货物数量为 $N=100$，假设其中 $M=15$ 个货物是次品。现抽取 $n=5$ 个样本，其中 x 个样本是次品的概率由如下表达式给出：

$$P(X=x)=\frac{\binom{15}{x}\binom{100-15}{5-x}}{\binom{100}{5}} \tag{3-305}$$

$x=0$，1，2，3，4 和 5 对应的概率分别是 0.4357，0.4034，0.1377，0.0216，0.0015 和 3.99×10^{-5}。◀

3.9.7 离散均匀随机变量

离散均匀随机变量是连续均匀随机变量的一个相似副本，它的概率密度函数为：

$$f_X(x)=\frac{1}{N}\sum_{n=a}^{b}\delta(x-n) \tag{3-306}$$

其中 a，$b\in\mathcal{Z}$，$a<b$，$N\triangleq b-a+1$ 表示输出结果的数量。累积分布函数为：

$$F_X(x)=\frac{1}{N}\sum_{n=a}^{b}u(x-n) \tag{3-307}$$

离散均匀概率密度函数的例子如图 3-48 所示。当 $a=1$，$b=N$ 时为标准离散均匀概率密度函数。

离散均匀分布的应用场合包括：

- 输出结果为等概率的试验，比如投掷一枚骰子。
- 数字通信系统中的脉冲幅度调制(PAM，见第 10 章)。

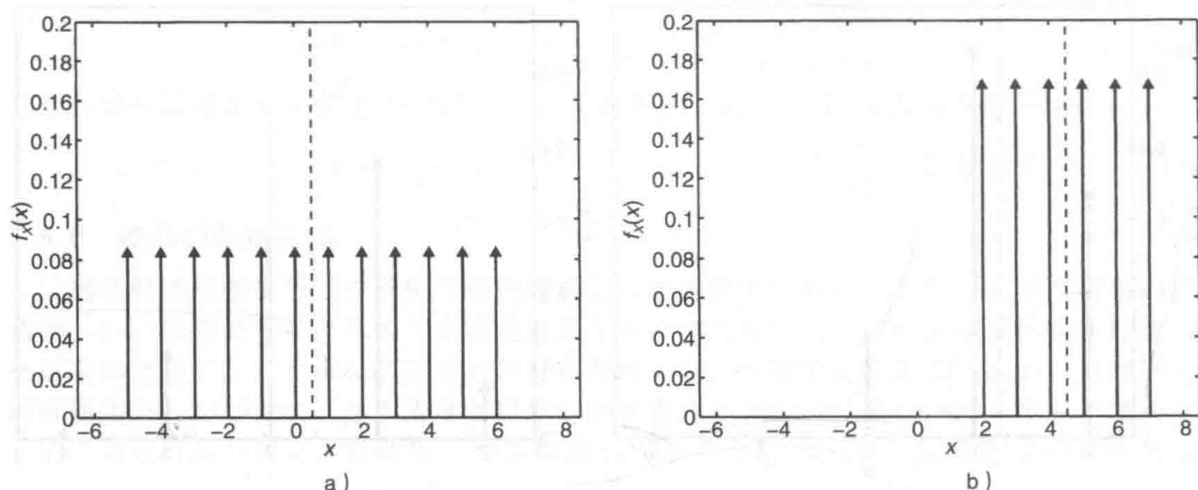


图 3-48 离散均匀分布概率密度函数的例子。a) $a=-5$，$b=6$；b) $a=2$，$b=7$。支集为 $\{a,\cdots,b\}$，虚线指示了 μ_X 的位置

例 3-66 离散均匀随机变量 X 的均值为：

$$\mu_X=\frac{1}{N}\int_{-\infty}^{\infty}\sum_{n=a}^{b}x\delta(x-n)\mathrm{d}x=\frac{1}{N}\sum_{n=a}^{b}n \tag{3-308}$$

可以用两种方式来写求和表达式：第一种按照 a 的升序，第二种按照 b 的降序：

$$\begin{aligned}\mu_X&=\frac{1}{N}[a+(a+1)+\cdots+(a+N-1)]=\frac{1}{N}\Big(aN+\sum_{k=0}^{N-1}k\Big)\\&=\frac{1}{N}[(b-N+1)+\cdots+(b-1)+b]=\frac{1}{N}\Big(bN-\sum_{k=0}^{N-1}k\Big)\end{aligned} \tag{3-309}$$

因为上面两个表达式的最后一行肯定是相等的，可以将二者相加再除以 2 得到最终结果：$\mu_X=(a+b)/2$。有限级数求和的结论也可以用来计算均值(见附录 E)：

$$\sum_{n=1}^{N}n=N(N+1)/2\Rightarrow\sum_{n=0}^{N-1}n=N(N-1)/2 \tag{3-310}$$

进行变量替换 $m=n-a$ 后：

$$\frac{1}{N}\sum_{m=0}^{b-a}(m+a)=\frac{1}{N}\Big[a(b-a+1)+\sum_{m=0}^{b-a}m\Big] \tag{3-311}$$

因此可得：

$$(1/N)[aN+(b-a+1)(b-a)/2]=(a+b)/2 \tag{3-312}$$

采用类似的方法可计算出方差为(见习题 5-36)：

$$\sigma_X^2=(N^2-1)/12 \tag{3-313}$$

请注意，尽管$\{a, b\}$和 N 在这里必须取整数，但离散均匀随机变量的均值和方差与连续情况的结果是非常相似的。◀

3.9.8 对数随机变量(对数级数)

对数随机变量的概率密度函数为：

$$f_X(x)=\frac{-1}{\ln(q)}\sum_{n=1}^{\infty}(p^n/n)\delta(x-n) \tag{3-314}$$

其中 $p\in[0, 1]$，$q\triangleq1-p$。因为 $\ln(q)$是负数，因此上式中有一个负号，这样确保 $f_X(x)$是一个合法的概率密度函数。对数分布(也称作对数级数分布)来自于下面的迈克劳林级数：

$$\ln(1-p)=-\sum_{n=1}^{\infty}p^n/n \tag{3-315}$$

公式两边同时除以 $\ln(1-p)=\ln(q)$，可以看出式(3-314)的求和结果为 1。对数概率密度函数的例子如图 3-49 所示。累积分布函数为：

$$F_X(x)=\frac{-1}{\ln(q)}\sum_{n=1}^{\infty}(p^n/n)u(x-n) \tag{3-316}$$

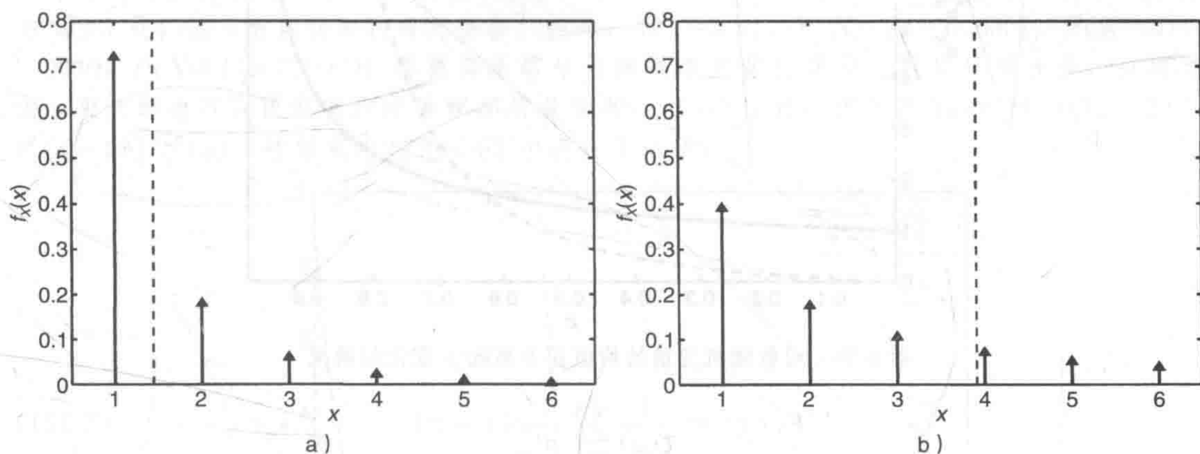


图 3-49 对数概率密度函数的例子。a) $p=0.5$；b) $p=0.9$。支集为$\mathcal{N}$，虚线指示了 μ_X 的位置

对数分布的应用场合包括：

- 对一个地区某物种丰富度的建模。
- 一个消费者在某段时间内购买的商品数量的分布。

例 3-67 对于如下概率质量函数的比值：

$$\frac{p_X[x+1]}{p_X[x]}=\frac{x}{x+1}p \tag{3-317}$$

可以看出对数概率质量函数是严格递减的。(在附录 E 中对离散随机变量的函数表达式进行了总结归纳。)因此，对数分布的众数必然是 $m_o=1$。均值为：

$$\begin{aligned}\mu_X&=\frac{-1}{\ln(q)}\int_{-\infty}^{\infty}\sum_{n=1}^{\infty}x(p^n/n)\delta(x-n)\mathrm{d}x=\frac{-1}{\ln(q)}\sum_{n=1}^{\infty}p^n\\&=\frac{-1}{\ln(q)}[1/(1-p)-1]=\frac{-1}{\ln(q)}(p/q)\end{aligned} \tag{3-318}$$

从图 3-50 可以看出，当 $p\to1$ 时 μ_X 会急剧的增大，这说明当 p 接近于 1 时概率密度函数的拖尾会非常长，这与图 3-49 不同(p 相对较小)。长拖尾的原因在于：当 $q\to0$ 时，概率密度函数分母中的 $\ln(q)$接近于 0。方差为 $\sigma_X^2=-p[p+\ln(q)]/q^2\ln^2(q)$，结果如图 3-50所示(见习题 5-36)。 ◀

3.9.9 泽塔随机变量(Zipf)

泽塔(Zeta)随机变量是基于黎曼泽塔函数得到的：

$$\zeta(\alpha)\triangleq\sum_{n=1}^{\infty}\frac{1}{n^\alpha} \tag{3-319}$$

其中 $\alpha>1$ 是一个实数值(复数也可以定义泽塔函数，此时要求 $\mathrm{Re}(\alpha)>1$)。请注意，$\zeta(\alpha)$是附录 B 中讨论过的幂级数。泽塔分布的概率密度函数为：

$$f_X(x)=\frac{1}{\zeta(\alpha)}\sum_{n=1}^{\infty}\frac{1}{n^\alpha}\delta(x-n) \tag{3-320}$$

其中 $\zeta(\alpha)$是一个归一化因子，这样确保概率密度函数的积分为 1。累积分布函数为：

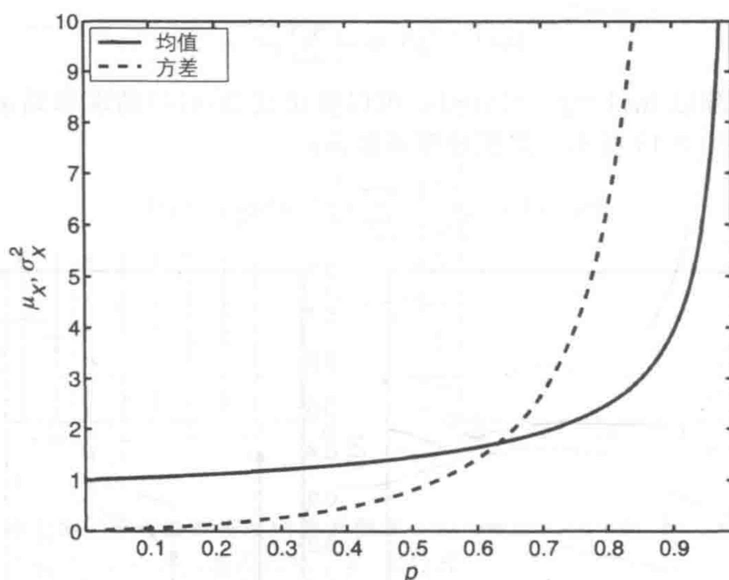


图 3-50　对数随机变量的均值和方差随 p 变化的情况

$$F_X(x)=\frac{1}{\zeta(\alpha)}\sum_{n=1}^{\infty}\frac{1}{n^{\alpha}}u(x-n) \tag{3-321}$$

泽塔分布的概率密度函数例子见图 3-51。可以看出，$P(X=1)=F_X(1)=1/\zeta(\alpha)$随着$\alpha$的增长而快速增大，所以概率质量函数被输出结果 $X=1$“统治”了。当 $N\to\infty$时，泽塔分布可以被看做 Zipf 分布的一种特例(也称作 Zipf 定律)。Zipf 分布的概率密度函数为：

$$f_Y(y)=\frac{1}{H_N(\alpha)}\sum_{n=1}^{N}\frac{1}{n^{\alpha}}\delta(y-n) \tag{3-322}$$

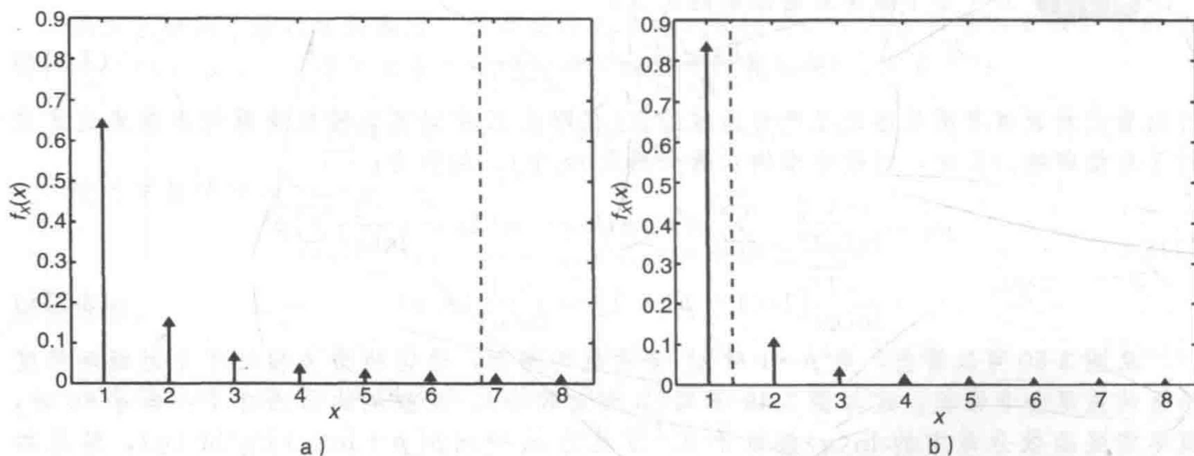


图 3-51　泽塔概率密度函数的例子。a) $\alpha=2.1$；b) $\alpha=3$。支集为$\mathcal{N}$，虚线指示了 μ_X 的位置

式(3-323)表示第 n 个广义调和数(见附录 B)。

$$H_n(\alpha)\triangleq\sum_{m=1}^{n}\frac{1}{m^{\alpha}} \tag{3-323}$$

泽塔分布的应用场合包括：

- 在一本厚书中单词出现频率的建模。
- 对绝大部分输出结果都很少发生的情况的建模。

例 3-68　与对数分布类似，泽塔分布的概率质量函数也是严格递减的：

$$\frac{p_X[x+1]}{p_X[x]}=\left(\frac{x}{x+1}\right)^{\alpha} \tag{3-324}$$

因此，众数一定为 $m_o=1$。均值比较容易计算：

$$\mu_X=\frac{1}{\zeta(\alpha)}\sum_{n=1}^{\infty}\frac{1}{n^{\alpha}}\int_{-\infty}^{\infty}x\delta(x-n)\mathrm{d}x=\frac{1}{\zeta(\alpha)}\sum_{n=1}^{\infty}\frac{n}{n^{\alpha}}=\frac{\zeta(\alpha-1)}{\zeta(\alpha)} \tag{3-325}$$

只有当 $\alpha-1>1\Rightarrow\alpha>2$ 时上面的无限求和才是收敛的。从图 3-52 可以看出，当 α 接近 2 的时候均值变得非常大，这说明在拖尾部分的概率质量很大(也就是说有很长的拖尾)。当 α 增大时，均值趋向于 1，这说明拖尾部分的概率变得非常小，几乎所有的概率都集中在 $X=1$ 处，这与前面提到的情况吻合。例如，当 $\alpha=5$ 时，$P(X=2)=0.0301$，$P(X=3)=0.0040$，$P(X=4)=0.0009$。需要将泽塔分布的概率质量函数用对数坐标来表示，这样即使 α 很大时也可以很方便的观察概率质量函数。当 $\alpha>3$ 时，方差为 $\sigma_X^2=[\zeta(\alpha)\zeta(\alpha-2)-\zeta^2(\alpha-1)]/\zeta^2(\alpha)$，该结果也在图 3-52 中进行了体现。

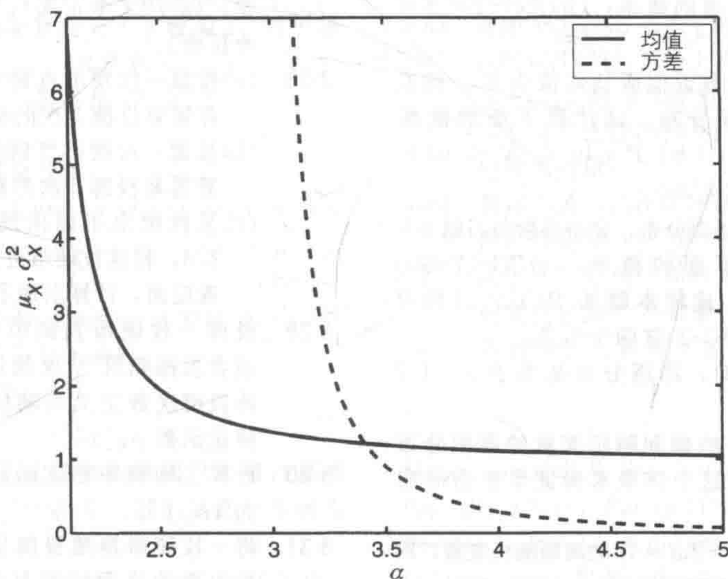


图 3-52 泽塔随机变量的均值和方差随 α 变化的情况

习题

随机变量及其分布

3-1 设 $\Omega=\{-3,-2,-1,0,1,2,3\}$，计算下面每个随机变量的映射结果：(a) $X_1(\omega)=\mathrm{sgn}(\omega)$；(b) $X_2(\omega)=\omega^2$；(c) $X_3(\omega)=I_{\{-3,-1,1,3\}}[\omega]+2I_{\{-2,0,2\}}[\omega]$。如果 Ω 的各个输出结果都是等概率的，计算这三个随机变量的累积分布函数 $F_X(x)$。

3-2 设 $\Omega=\{1,\cdots,10\}$，重复计算习题 3-1，随机变量为：(a) $X_1(\omega)=u(\omega-4.5)$；(b) $X_2(\omega)=(10-\omega/2)I_{\{2,4,6,8,10\}}[\omega]$；(c) $X_3(\omega)=\mathrm{sgn}(\omega-2)$。

3-3 设 $\Omega=[-1,1]$，请重复计算习题 3-1，随机变量为：(a) $X_1(\omega)=\exp(\omega)$；(b) $X_2(\omega)=\cos(\omega\pi/2)$；(c) $X_3(\omega)=\mathrm{sgn}(\omega)$。为了计算 $F_X(x)$，在此假设 Ω 的输出服从均匀分布。

3-4 设 Ω 为投掷骰子试验的样本空间，定义 $X(\omega)=I_{\{1,2\}}(\omega)+2I_{\{3,4,5\}}(\omega)-I_{\{6\}}(\omega)$。给出 Ω 最小的 σ-域，使得 $X(\omega)$ 为一个随机变量，并计算每个输出结果的概率值。

3-5 对于下面这些样本空间，给出最小的 σ-域，使得 $X(\omega)=\lfloor\omega\rfloor$ 为一个随机变量。(a) $\Omega_1=\mathcal{R}^+$；(b) $\Omega_2=[0,4)$。

3-6 验证下面这些随机变量 X 的分布函数是否合法：(a) $\sum_{n=1}^{\infty}p^n u(x-n)$，$x\in\mathcal{Z}$，$0<p<1$；(b) $x[u(x)-u(x-2)]$，$x\in\mathcal{R}$；(c) $(1/2)[u(x)+\exp(x-1)u(x-1)]$，$x\in\mathcal{R}$。

3-7 (a)计算并绘出下面概率密度函数的累积分布函数：

$$f_X(x)=(1/4)[u(x-1)-u(x+1)]+(1/6)\delta(x-1)+(1/3)\delta(x-2) \tag{3-326}$$

(b)重复(a)的要求：

$$f_X(x)=\begin{cases}0, & x<-1\\ x+1, & -1\leqslant x<0\\ (1/2)\exp(-x), & x\geqslant 0\end{cases} \tag{3-327}$$

3-8 对于 $\Omega=[-1,2]$ 和 $X(\omega)=\omega^2-1$，假设 Ω 上的概率测度为乘以 1/3 后的 L 测度，计算 $F_X(x)$ 和 $f_X(x)$ 的表达式。

连续随机变量

3-9 对于分布参数为 $\lambda=2$ 的指数随机变量 X，试计算下面这些概率。(a) $P(1<X<3)$；(b) $P(X>5)$；(c) $P(X>5\mid X>3)$；(d) $P(X>2\mid X>3)$。

3-10 设 X 是参数为 $\{\mu,\sigma\}$ 的高斯随机变量。(a)用标准高斯随机变量的累积分布函数 $F_o(x)$ 来表示 $P(x_1\leqslant X\leqslant x_2)$。如果 $\mu=0$，$\sigma=2$，计算下面的概率：(b) $P(-2<X<1)$；(c) $P(|X|>2)$。

3-11 某次考试的分数近似服从均值为 60，标准差为 5 的高斯分布。试计算下面的概率 (a) $P(X>90)$；(b) $P(40<X<65)$；(c) $P(|X-\mu|\leqslant 2\sigma)$。

3-12 设 Y 服从对数正态分布，分布参数与习题 3-10 相同。计算下面的概率：(a) $P(1<Y<3)$；(b) $P(Y>2)$。比较本题的 $P(Y>2)$ 和习题 3-10的 $P(X>2)$ 这两个结果。

3-13 证明当 $\alpha=\beta$ 时，泽塔分布是关于 $x=1/2$ 对称的。

3-14 推导参数为 α 的瑞利随机变量的累积分布函数，并请用这个结果来验证半正态分布的均值。

3-15 对于参数为 $\{\mu=2,\sigma=1\}$ 的高斯随机变量，用 Q 函数来表示下面的概率：(a) $P(|X|>\sigma)$；(b) $P(-3\leqslant X-\mu\leqslant 1)$。再用误差函数来表示这些结果。

3-16 推导 F 分布众数的表达式，即式(3-224)。

3-17 X 为参数为 $\alpha=2$ 的瑞利随机变量，计算：(a) $P(1\leqslant X\leqslant 3)$；(b) $P(X\leqslant 4$ 或 $X>3)$；(c) $P(X>4$ 或 $X\leqslant 3)$。

3-18 证明式(3-187)给出的 Logistic 随机变量的累积分布函数 $F_X(x)$。

3-19 推导对称拉普拉斯分布的中位数 m_e 表达式(式(3-104))，分布参数为 $\{\alpha_1,\alpha_2,c\}$。

3-20 以 $x=a\pm b$ 的形式给出高斯随机分布的拐点，并将计算结果与例 3-25 中柯西随机变量的拐点相比较。

3-21 推导三角分布的中位数 m_e 表达式(式(3-122))，分布参数为 $\{a,b,c\}$。

3-22 如果 X 服从泽塔分布，分布参数为 $\{\alpha=r,\beta\}$，当 $\beta\to\infty$ 时，证明 $Y=\beta X$ 的极限分布为伽马分布，且分布参数为 $\{\lambda=1,r\}$。

3-23 设 X 服从区间 $[0,10]$ 上的均匀分布。假设在该区间上随机的取一个值 x。将该区间分为两个子区间。假设 Y 为较长的子区间的长度，计算随机变量 Y 的概率密度函数。

3-24 对卡方分布随机变量进行转换，对于参数为 α 的麦克斯韦-玻耳兹曼分布，证明其均值为 $\mu_X=2\alpha\sqrt{2/\pi}$。

3-25 可根据变换 $X+c$ 得到三参量的韦伯分布，其中 $c\geqslant 0$。计算该平移分布的中位数，众数以及任意一个拐点。

离散随机变量

3-26 对于分布参数为 $\alpha=2$ 的泊松随机变量 X，计算下面这些概率：(a) $P(X>3)$；(b) $P(2<X\leqslant 5)$；(c) $P(X=3\mid 2<X\leqslant 5)$。

3-27 将一枚硬币投掷连续 5 次，计算下面的概率：(a) P(3 个正面)；(b) P(0 个正面或 0 个反面)。

3-28 (a)投掷一枚硬币直到第一次出现正面，计算需要投掷 3 次的概率。

(b)投掷一枚硬币直到第二次出现正面，计算需要投掷 5 次的概率。

(c)某枚硬币正面出现的概率为 $P(H)=2/3$，将这枚硬币进行投掷直到第三次出现反面，计算出现了 5 次正面的概率。

3-29 投掷一枚硬币直到第一次出现正面(H)，或者投掷到第 N 次就停止。设 $P(H)=p$，将投掷次数定义为随机变量 X，计算概率质量函数 $p_X[x]$。

3-30 计算二项概率密度函数的众数，分布参数为 $\{N,p\}$。

3-31 将一枚硬币连续投掷 4 次。将第一次正面后出现的反面的数量定义为随机变量 X，计算概率密度函数 $f_X(x)$。

3-32 对于一个二元对称信道，输入为 X，输出为 Y，符号表为 $\{0,1\}$，信道错误概率为 $P(Y=0\mid X=1)=P(Y=1\mid X=0)=0.1$。为了提高传输的准确性，我们将每个比特重复两次传输：发送 000 表示 0，发送 111 表示 1。如果采用这种发送方法，计算此时符号错误的概率。

3-33 两个人轮流投掷一枚硬币直到第二次出现正面。我们直观的感觉是认为先投掷的那个人获胜的概率会较高。计算每个人的获胜概率，并据此判断这个直观感觉是否正确。

3-34 对于二项随机变量 X，分布参数为 $\{p=1/4,N=5\}$，计算：(a) $P(X\leqslant 2)$；(b) $P(X\leqslant 2$ 或 $X>3)$；(c)计算 x 使之满足 $P(X>x)<0.01$。

3-35 假设一个教授拥有 15 本概率论的书，为方便起见将每本书进行编号。一个研究生每天随机的借走两本书，并在当天晚上归还。请计算下面这些概率，并确定离散分布的类型。(a)在前 5 天，第 4 本书被借走两次的概率；(b)直到第 3 天，第 1 本书都没被

借走的概率。

3-36 如果我们对几何随机变量的概率 $P(X>k)$ 感兴趣，对于 $0<\epsilon<1$ 有：

$$\int_{k+\epsilon}^{\infty} f_X(x)\mathrm{d}x = \sum_{n=k+1}^{\infty} pq^n \quad (3\text{-}328)$$

因为 $X>k$ 因此 $n=k$ 的情况被排除在外。请给出上面求和结果的闭合表达式。

3-37 推导二项分布概率的递归表达式，即式(3-252)。

3-38 假设一个离散随机变量具有性质 $P(X=x+1)=aP(X=x)$，$x\in\mathcal{Z}^+$。给出 $p_X[x]$的表达式，并证明 $1/2<P(X$ 是偶数$)<1$。

3-39 已知泊松随机变量的分布参数为 α，对于非整数的 $x\in\lfloor\alpha\rfloor,\infty)$和整数 $x\in[\alpha-1,\infty)$，证明概率密度函数都是非递增的。

混合随机变量

3-40 计算如下的混合分布：

$$F_X(x)=1/4+(1/6)u(x+1)+(1/3)u(x)+(1/4)\operatorname{sgn}(x-\mu)[1-\exp(-\alpha|x-\mu|)] \quad (3\text{-}329)$$

计算它的概率密度函数，并验证 $F_X(x)$是一个合法的累积分布函数。

3-41 对于下面的概率密度函数：

$$f_X(x)=(1/8)\delta(x-1)+(1/8)\delta(x-2)+(1/4)(u(x-1.5)-u(x-2.5)]+(1/2)\exp(-x)u(x) \quad (3\text{-}330)$$

(a)计算 $P(X>1)$和 $P(X\leqslant 2)$；(b)给出累积分布函数 $F_X(x)$的表达式。

仿真作业

3-42 请使用 rand 函数来模拟习题 3-23。产生 Y 的 1000 个样本并绘出直方图，观察该直方图与理论推导的概率密度函数是否吻合。

3-43 使用 MATLAB 来产生例 3-44 的结果(图 3-34)。同时将区间 $[-1,1]$ 上均匀随机变量以及 $c=(a+b)/2$ 的三角随机变量的对应结果放在一起进行比较(均匀分布和三角分布都具有轻微的拖尾)。

4-44 利用 MATLAB 从分布参数为$\{\mu=2,\sigma=1\}$的高斯分布中产生 100 个样本，根据学生氏 t 分布分别计算样本均值为 90%和 95%的置信区间。再根据卡方分布来计算样本方差为 90%和 95%的置信区间。

进一步阅读

附录 A 中提供了参数连续和离散随机变量许多参考文献。在下面的这些参考文献中，进一步研究了概率论和随机变量的有关知识。

Davenport，Jr. (1970)，Gray 和 Davisson(2004)，Larson 和 Shubert(1979)，Leon-Garcia(2008)，Mood、Graybill 和 Boes(1974)，Olofsson(2005)，Papoulis(1965)，Parzen(1960)，Peebles(1993)，Peebles(1993)，Robinson(1985)，Shiryayev(1984)，Stark 和 Woods(2002)，Viniotis(1998)。

第4章 多维随机变量

4.1 引言

本章目的是将一维随机变量的定义扩展到二维或多维联合随机变量并且研究它们之间的联系。首先对第3章学习的一些随机变量进行近似。这些近似不仅告诉我们这些随机变量如何近似，而且可以用来简化那些用原始随机变量的概率密度函数(pdf)比较复杂的计算。接下来，介绍用于推导事件概率的多维随机变量的联合概率分布，多维随机变量也称为随机向量。这里将学习条件分布，即基于条件事件或条件随机变量来计算概率。图4-1a给出了在特定值 x 下的条件概率密度函数 $f_{Y|X}(y|x)$。条件是非常重要的，因为它描述了两个随机变量的相关程度以及它们的“互信息”。

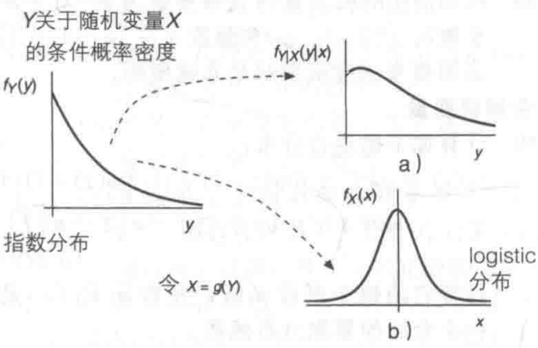


图 4-1　随机变量的变换。a) 以相关随机变量 X 为条件对应一个特定值 x 时的 $f_{Y|X}(y|x)$；b) 变换为一个新的随机变量 X，概率密度函数为 $f_X(x)$

本章还将学习独立的概念，它将在第6、7章中来描述随机过程的一个特性。借由几个例子，我们将学习怎样通过其他随机变量的变换推导随机变量的概率密度函数和累积分布函数。例如，指数分布的随机变量经过特定的非线性变换后变成了Logistic随机变量，如图4-1b所示。另外，本章研究的变换中还包含随机变量的一个重要集合——顺序统计量。最后，我们将讨论混合的概念，它不是随机变量的变换，而是它们概率密度函数的加权组合。

4.2 随机变量的近似

本节介绍的近似不是随机变量概率密度函数中的参数为特定有限值的特殊情况，而是一个或多个参数非常大时的极限情况。表4-1总结了一些相关的离散随机变量，我们将首先对此展开讨论。

表 4-1　随机变量的渐进近似

分布	条件	近似
超几何分布 $\{N,M,n\}$	$N,\ M\to\infty,\ M/N=p$	二项分布 $\{n,\ p\}$
二项分布 $\{n,p\}$	$n\to\infty,\ p\to 0,\ np=\alpha$	泊松分布 $\{\alpha\}$
负二项分布 $\{m,p\}$	$m\to\infty$	泊松分布 $\{m(1-p)\}$
二项分布 $\{N,p\}$	N 较大	高斯分布 $\{\mu=Np,\ \sigma^2=Npq\}$
超几何分布 $\{N,M,n\}$	$N,\ M\to\infty$，M/N 为常数	高斯分布 $\{\mu=nM/N,\sigma^2=n(M/N)(1-M/N)\}$
泊松分布 $\{N,\alpha\}$	N 较大	高斯分布 $\{\mu=N\alpha,\sigma^2=N\alpha\}$
卡方分布 $\{N\}$	N 较大	高斯分布 $\{\mu=N,\sigma^2=2N\}$
学生氏 t 分布 $\{r\}$	r 较大	高斯分布 $\{\mu=0,\sigma^2=1\}$
伽马分布 $\{\lambda,r\}$	r 较大	高斯分布 $\{\mu=r/\lambda,\sigma^2=r/\lambda^2\}$

4.2.1　超几何分布的二项分布近似

考虑超几何分布的概率质量函数，参数 $N\in\mathcal{N}, M\in\{0,\cdots,N\}, n\in\{1,\cdots,N\}$，代入 $M=Np$：

$$P(X=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}=\frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}} \tag{4-1}$$

令 $M\to\infty$，$N\to\infty$，并且二者比值 $P\triangleq M/N$ 为常数。展开二项式系数，组合一些项得到 $\binom{n}{x}$，消去分子、分母的公共项，得：

$$\begin{aligned}&\frac{(Np)!}{x!(Np-x)!}\frac{(N-Np)!}{(n-x)!(N-Np-n+x)!}\frac{n!(N-n)!}{N!}\\&=\binom{n}{x}\prod_{k=0}^{x-1}(Np-k)\prod_{m=0}^{n-x-1}(N-Np-m)\frac{1}{\prod_{k=0}^{n-1}(N-k)}\end{aligned} \tag{4-2}$$

式中，分子有 $x+(n-x)=n$ 项（不包括前面的 $\binom{n}{x}$ 因子），分母也有相同数目的项。因此，可以将分母比照分子乘积的形式分为两部分（注意 $x\leqslant n$）：

$$\binom{n}{x}\prod_{k=0}^{x-1}\frac{(Np-k)}{(N-k)}\prod_{m=0}^{n-x-1}\frac{(N-Np-m)}{(N-x-m)}\xrightarrow{N\to\infty}\binom{n}{x}\prod_{k=0}^{x-1}p\prod_{m=0}^{n-x-1}q=\binom{n}{x}p^x q^{n-x} \tag{4-3}$$

可得：

$$\lim_{N\to\infty}\frac{Np-k}{N-k}=\lim_{N\to\infty}\frac{Np}{N}=p \tag{4-4}$$

$$\lim_{N\to\infty}\frac{N-Np-k}{N-x-k}=\lim_{N\to\infty}\frac{N-Np}{N}=1-p\triangleq q \tag{4-5}$$

式(4-3)的最后一个表达式是参数为 $\{n,p\}$ 的二项分布随机变量的概率质量函数。图 4-2给出了 $p=M/N=16/24$，$n=6$ 时超几何分布的二项分布近似。

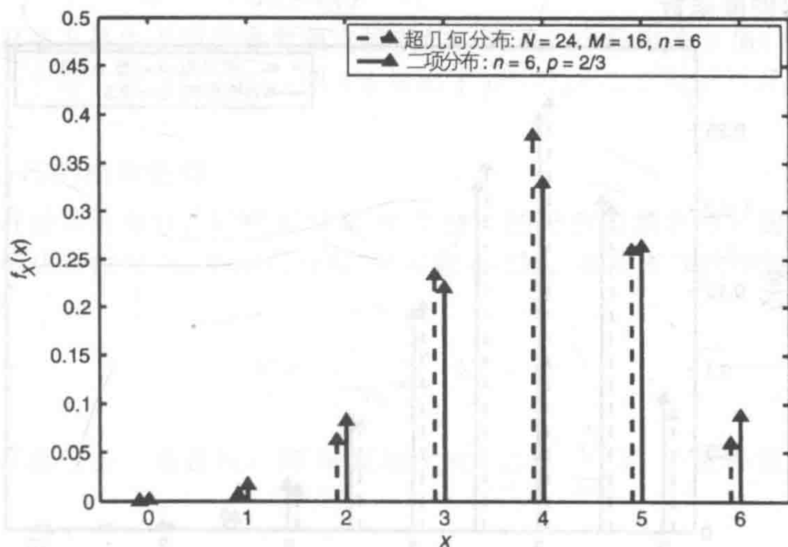


图 4-2　超几何分布随机变量的二项分布近似举例（为便于比较，超几何分布的 pdf 略向左平移）

4.2.2　二项分布的泊松近似

设随机变量服从参数为 $\{n,p\}$ 的二项分布，其中 $p\in[0,1]$ 为成功概率，$n\in\mathcal{Z}^+$ 为试

验次数。当p非常小而n非常大时，二项分布可由泊松分布近似。近似的关键是当 $n\to\infty$、$p\to 0$时 $np\triangleq\alpha>0$ 保持不变，其中 α 是泊松分布随机变量的参数。令 $p=\alpha/n$，$q\triangleq 1-p$，重新整理二项分布的概率质量函数：

$$\begin{aligned}\lim_{n\to\infty}\binom{n}{x}p^x q^{n-x} &=\lim_{n\to\infty}\binom{n}{x}(\alpha/n)^x(1-\alpha/n)^{n-x}\\ &=\lim_{n\to\infty}\frac{n!}{(n-x)!n^x}(\alpha^x/x!)(1-\alpha/n)^n(1-\alpha/n)^{-x}\end{aligned}\tag{4-6}$$

观察式(4-6)中包含 n 的三项，可得

$$\lim_{n\to\infty}(1-\alpha/n)^{-x}=1\tag{4-7}$$

$$\lim_{n\to\infty}\frac{n!}{(n-x)!n^x}=\lim_{n\to\infty}\frac{n(n-1)\cdots(n-x+1)}{n^x}=1\tag{4-8}$$

$$\lim_{n\to\infty}(1-\alpha/n)^n=\exp(-\alpha)\tag{4-9}$$

式(4-7)显而易见；式(4-8)是因为分子里 n 的最大次幂为 x，与分母相等；式(4-9)可以利用二项式定理得到(见附录 B)：

$$\begin{aligned}(1-\alpha/n)^n &=1+n(-\alpha/n)+n(n-1)(-\alpha/n)^2/2!+n(n-1)(n-2)(-\alpha/n)^3/3!\\ &\quad+\cdots+(-\alpha/n)^n,\\ &=1-\alpha+[(n-1)/n]\alpha^2/2!-[(n-1)(n-2)/n^2]\alpha^3/3!\\ &\quad+\cdots+(-\alpha/n)^n\end{aligned}\tag{4-10}$$

当 $n\to\infty$时，包含 n 的所有比值趋近于 1 并且$(1/n)^n\to 0$，因此：

$$\lim_{n\to\infty}(1-\alpha/n)^n=\sum_{m=0}^{\infty}(-\alpha)^m/m!\tag{4-11}$$

它是 $\exp(-\alpha)$的展开式，式(4-9)得证。将式(4-7)、式(4-8)、式(4-9)代入式(4-6)，得到参数为 α 的泊松分布：

$$\lim_{n\to\infty}\binom{n}{x}p^x q^{n-x}=\exp(-\alpha)\frac{\alpha^x}{x!}=P(X=x)\tag{4-12}$$

图 4-3 给出了 $p=0.1$，$n=25$ 的二项分布概率密度函数的近似，也就是 $\alpha=np=2.5$ 的泊松分布概率密度函数。

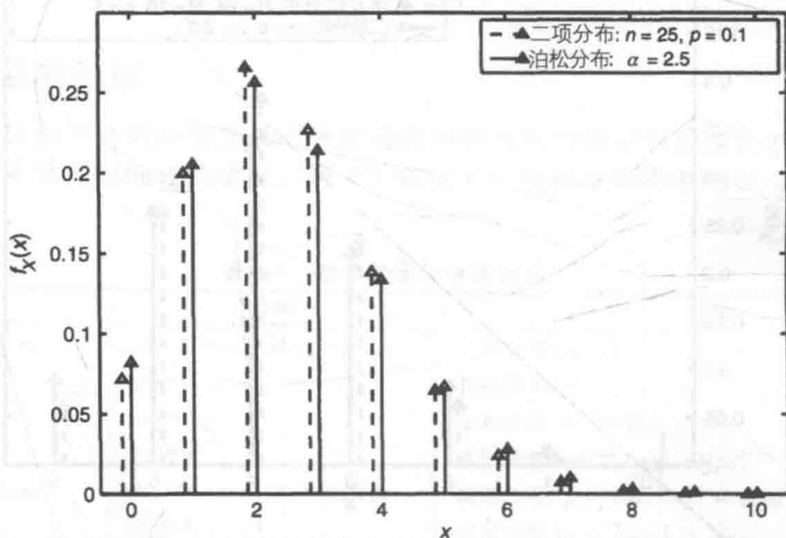


图 4-3　泊松分布随机变量的二项分布近似(为便于比较，二项分布的 pdf 略向左平移)

表 4-1 中的负二项分布的泊松分布近似留作习题 4-3。

4.2.3 高斯近似

表 4-1 总结了前面学习的离散随机变量可用高斯分布近似，这些近似都是用第 7 章的中心极限定理(Central Limit Theorem，CLT)推导出来的。表 4-1 给出的卡方分布、学生氏 t 分布的高斯分布近似留作本章习题。

定理 4-1(中心极限定理) 假设 $\{X_1,\cdots,X_N\}$ 为独立同分布(iid)的随机变量，其数学期望 μ_X 和方差 σ_X^2 为有限值，则

$$\lim_{N\to\infty} P\left(\frac{1}{\sigma_X\sqrt{N}}\sum_{n=1}^{N}(X_n-\mu_X)\leqslant z\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp(-x^2/2)\mathrm{d}x \tag{4-13}$$

式(4-13)是参数为 $\{\mu=0,\sigma=1\}$ 的标准高斯累积分布函数。

证明：见第 7 章。

观察发现，减去 μ_X 再除以 $\sqrt{N\sigma_X^2}$ 后使得和的均值为 0 且方差为 1。也就是说，如果“足够大”数量的独立同分布正态随机变量加在一起，和的分布近似服从标准高斯分布。为了用中心极限定理(CLT)推导离散随机变量的高斯近似，必须有足够多的随机变量相加。(独立随机变量在本章稍后定义，它类似于第 2 章介绍的独立事件。)

4.2.4 二项分布的高斯近似

二项分布随机变量可由成功概率为 p 的 N 个独立同分布伯努利随机变量$\{X_k\}$之和得到。其均值和方差分别为 Np 和 Npq。根据中心极限定理：

$$Y=\frac{1}{N\sqrt{pq}}\sum_{n=1}^{N}(X_n-Np) \tag{4-14}$$

近似服从标准高斯分布。可以将上式中的求和看成多个随机变量到一个随机变量的变换(本章稍后将介绍随机变量的变换)。或者当 N 足够大时，$Z=\sum_{n=1}^{N}X_n$ 的概率密度近似服从均值为 Np、方差为 Npq 的高斯分布：

$$f_Z(z)\approx\frac{1}{\sqrt{2\pi Npq}}\exp(-(z-Np)^2/2Npq) \tag{4-15}$$

注意，均值和方差不是由不同的参数独立控制的：对于一个给定的 p 值(或 q)，当 N 很大时，可以得到一个特定的高斯近似。图 4-4 给出了 $\mu=Np=2.5$ 和 $\sigma^2=Npq=2.25$ 时近似的例子。

4.2.5 泊松分布的高斯近似

与前面的近似方法类似，同样假设有 N 个独立同分布的参数为 α 的泊松分布随机变量，它们之和服从参数为 $N\alpha$ 的泊松分布(见习题 4-25)。如果求和式中每一项的均值和方差都是 α，那么

$$Y=\frac{1}{\sqrt{N\alpha}}\sum_{n=1}^{N}(X_n-\alpha)^2 \tag{4-16}$$

近似服从标准高斯分布。或者说，当 N 足够大时，$Z=\sum_{n=1}^{N}X_n$ 的概率密度函数可由下式近似：

$$f_Z(z)\approx\frac{1}{\sqrt{2\pi N\alpha}}\exp(-(z-N\alpha)^2/2N\alpha) \tag{4-17}$$

其均值为 $N\alpha$，方差为 $N\alpha$。与二项随机变量的高斯近似类似，均值和方差不是独立的由不同的参数决定，实际上二者相等。图 4-5 给出了 $\mu=\sigma^2=N\alpha=5$ 时的近似结果。

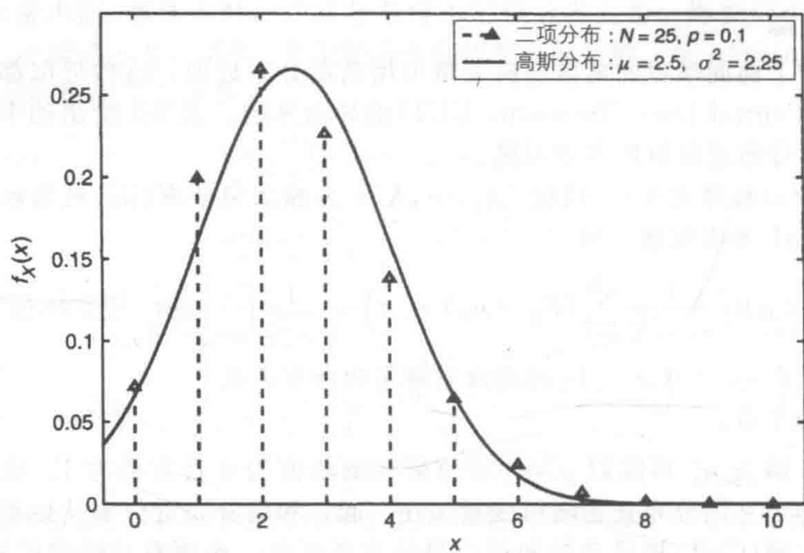


图 4-4 二项分布随机变量的高斯分布近似举例

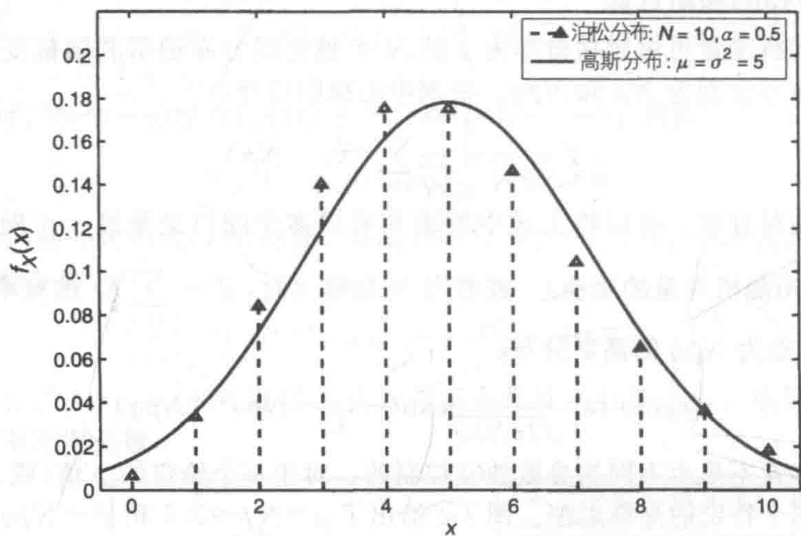


图 4-5 泊松分布随机变量的高斯分布近似举例

4.2.6 超几何分布的高斯近似

最后，由于二项分布随机变量可用高斯概率密度来近似，而前者又可以用来近似超几何分布，从而可以得到超几何分布随机变量的高斯近似。因此，同样假设 p 为常数，$\lim_{N\to\infty} M/N=p$，那么高斯近似的均值为 nM/N，方差为 $n(M/N)(1-M/N)$（利用了 $q\triangleq 1-p=1-M/N$）。同样，均值和方差不是独立的参数。图 4-6 给出了 $N=24$，$M=12$，$n=4$ 时这种分布的例子。

表 4-1 中的最后三个高斯近似留作本章习题，注意，这些近似不是基于表中第一列的随机变量之和；相反地，当随机变量的自由度较大时，这三个分布近似一个高斯分布。然而，考虑到第 3 章学过当自由度给定时，这些特定的随机变量可由其他随机变量相加得到，其中求和个数由自由度决定。因此，当卡方分布的 N、学生氏 t 分布和伽马分布的 r 参数很大时，中心极限定理同样适用。

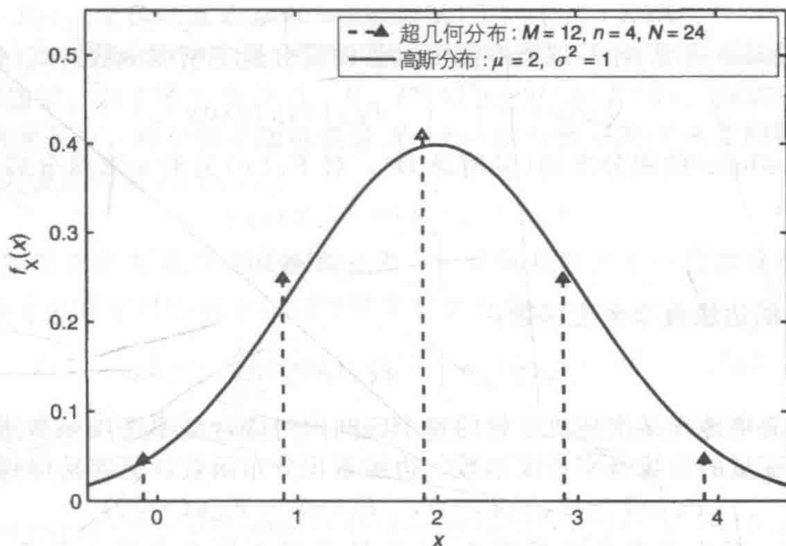


图 4-6 超几何分布随机变量的高斯分布近似举例

4.3 联合分布和边缘分布

我们希望得到多维随机变量的联合概率描述，从而量化它们的关系。在许多应用中这一点是非常重要的，如估计理论(见第 9 章和第 11 章)，就是从别的随机量的数据中估计或预测随机变量或随机过程的特性。第 3 章中学习了单个随机变量的概率密度函数，可以用类似的方法获得两个随机变量 X 和 Y 的联合概率密度函数。4.6 节介绍抽象的概率空间 $\{\Omega,\mathcal{F},\mathcal{P}\}$ 到多维随机变量的映射。这里仅给出联合概率密度函数的定义。

定义(联合概率密度函数) 用 $f_{X,Y}(x,y)$ 表示随机变量 X 和 Y 的联合概率密度函数，计算如下：

$$P(x_1 \leqslant X \leqslant x_2, y_1 \leqslant Y \leqslant y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x \tag{4-18}$$

其概率测度定义在观测空间 $\{\mathcal{R}^2, \mathcal{B}(\mathcal{R}^2)\}$ 上，而且关于 $\mathcal{R}^2$ 上的勒贝格测度完全连续。

以上概率通过在笛卡儿积 $[x_1,x_2]\times[y_1,y_2]$ 中对联合概率密度积分得到，$[x_1,x_2]\times[y_1,y_2]$ 是 $\mathcal{R}^2$ 上一个矩形。图 4-7a 给出了 $(-\infty,x]\times(-\infty,y]$ 上的交集。

利用 $f_{X,Y}(x,y)$ 可以得到 X 和 Y 各自的概率密度函数，在这里称边缘概率密度函数。从图 4-7b 可以看出 $\{X\leqslant x\}\cap\{-\infty\leqslant Y\leqslant\infty\}=\{X\leqslant x\}$，因此

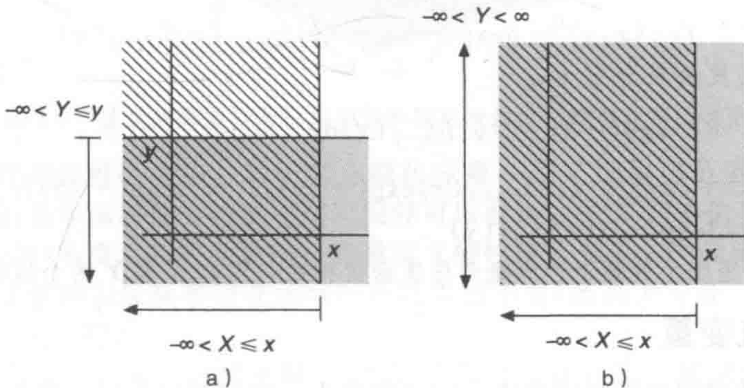


图 4-7 联合概率分布。a) 利用联合概率密度函数在 $(-\infty,x]\times(-\infty,y]$ 上的积分计算 $P(X\leqslant x, Y\leqslant y)$；b) 利用联合概率密度函数在 $(-\infty,x]\times(-\infty,+\infty]$ 上的积分计算 $P(X\leqslant x)$，延伸到 x-y 轴端点的矩形边缘为无穷大

$$P(X \leqslant x) = P(X \leqslant x, -\infty \leqslant Y \leqslant \infty) \tag{4-19}$$

确定事件$-\infty \leqslant Y \leqslant \infty$不影响 X 的概率测度。借助联合概率密度函数，式(4-19)等于

$$F_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{4-20}$$

利用莱布尼茨(Leibniz)的积分准则(见附录 D)，对 $F_X(x)$关于 x 求微分后，可得 X 的边缘概率密度函数：

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y \tag{4-21}$$

同理可得 Y 的边缘概率密度函数：

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x \tag{4-22}$$

总之，通过简单地在某个随机变量的整个区间内对联合概率密度函数进行积分，可以得到另一个随机变量的边缘概率密度函数。边缘累积分布函数就更容易得到了：

$$F_X(x) = F_{X,Y}(x,\infty), \quad F_Y(y) = F_{X,Y}(\infty,y) \tag{4-23}$$

例 4-1 图 4-8 给出了高斯随机变量 X 和 Y 的联合概率密度的例子。参数 $\mu_X=1$，$\mu_Y=0$，$\sigma_X^2=2$，$\sigma_Y^2=1$，相关系数 $\rho=0.9$(见第 5 章)。本章稍后将讨论此二维高斯概率密度函数以及图中 X 和 Y 对应的边缘概率密度函数。

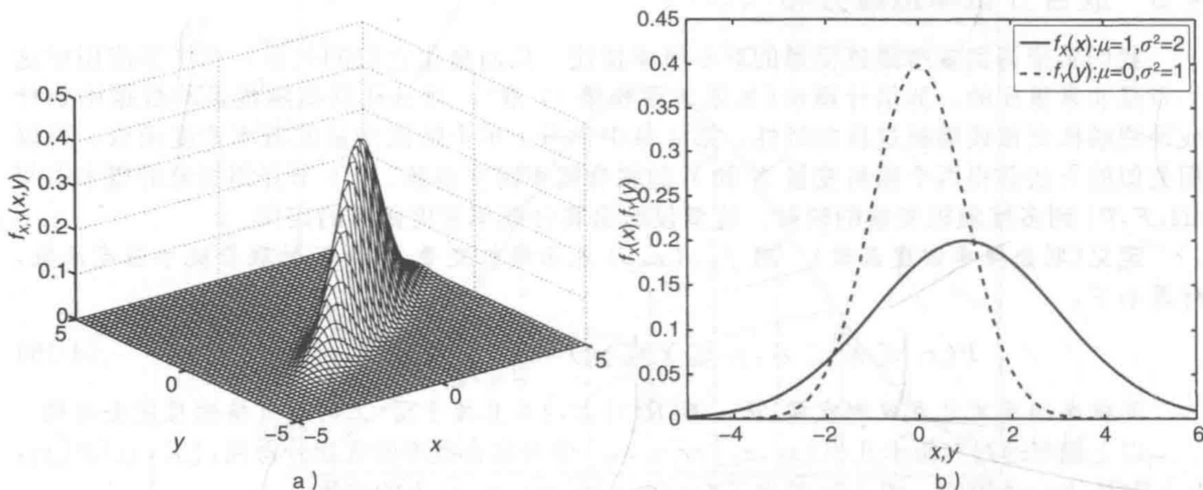


图 4-8　高斯随机变量。a) $\rho=0.9$ 时的二维概率密度函数；b) 边缘概率密度函数 ◀

例 4-2 下式中的联合概率密度函数：

$$f_{X,Y}(x,y) = 6\exp(-2x-3y)I_{[0,\infty)}(x)I_{[0,\infty)}(y) \tag{4-24}$$

对应的边缘概率密度函数分别为：

$$\begin{aligned} f_X(x) &= 6\int_0^{\infty} \exp(-2x-3y)I_{[0,\infty)}(x)\mathrm{d}y \\ &= -2\exp(-2x)I_{[0,\infty)}(x)\exp(-3y)\big|_0^{\infty} = 2\exp(-2x)I_{[0,\infty)}(x) \end{aligned} \tag{4-25}$$

$$f_Y(y) = 3\exp(-3x)I_{[0,\infty)}(y) \tag{4-26}$$

由于联合概率密度函数等于边缘概率密度函数相乘，因此 X 和 Y 两个随机变量独立。 ◀

4.4　独立随机变量

独立随机变量出现在许多应用中，它经常作为一个近似以导出一些问题的结果，否则这些结果将是不可能得到的。

定义(独立随机变量)　随机变量 X 和 Y 是独立的，当且仅当

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \tag{4-27}$$

其中，$F_X(x)$和$F_Y(y)$是边缘累积分布函数。

由于累积分布函数是事件发生的概率，因此上述定义不仅符合直观认识，而且符合第2章学习的概率知识：对于独立事件 A、B，$P(AB)=P(A)P(B)$。由式(4-27)可以直接得出联合概率密度函数，对于独立随机变量 X、Y，联合概率密度函数同样可以分解为乘积的形式(同样为独立的定义)：

$$f_{X,Y}(x,y)=f_X(x)f_Y(y) \tag{4-28}$$

例 4-3 设随机变量 X 服从标准高斯分布，Y 服从参数为 λ 的指数分布。如果它们是独立的，那么马上可得它们的联合(二维)概率密度函数：

$$f_{X,Y}(x,y)=\frac{\lambda}{\sqrt{2\pi}}\exp(-x^2/2-\lambda y)I_{[0,\infty)}(y)I_{(-\infty,\infty)}(x) \tag{4-29}$$

这可以扩展到更多的独立随机变量。如果 Z 在$[-1,1]$服从均匀分布，那么三个变量的联合概率密度为：

$$f_{X,Y,Z}(x,y,z)=\frac{\lambda}{2\sqrt{2\pi}}\exp(-x^2/2-\lambda y)I_{[0,\infty)}(y)I_{(-\infty,\infty)}(x)I_{[-1,1]}(z) \tag{4-30}$$

与式(4-29)相比，式(4-30)有 1/2 的系数，而且包含指示函数，该函数确定 Z 的支集。◀

例 4-4 设随机变量 X 服从标准高斯分布，Y 在$[-2,2]$上服从均匀分布。如果两个随机变量独立，那么它们的联合概率密度为：

$$f_{X,Y}(x,y)=\frac{1}{4\sqrt{2\pi}}\exp(-x^2/2)I_{(-\infty,\infty)}(x)I_{[-2,2]}(y) \tag{4-31}$$

该指示函数清楚地定义了取值范围，图 4-9 所示为该函数的三维图形。

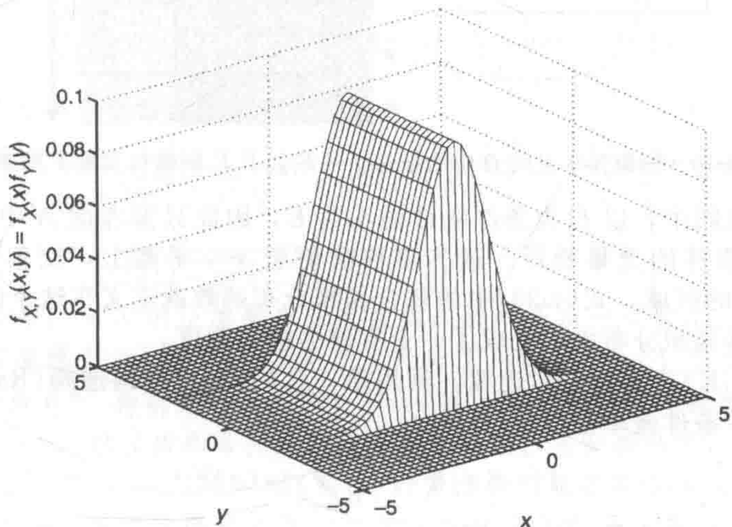


图 4-9　独立的高斯分布、均匀分布随机变量的二维概率密度函数 ◀

从函数的角度我们可以看出，对于独立随机变量，在 Y 的取值范围内，随机变量 X 的概率密度函数的基本形状保持钟形不变。同样可以看出，在 X 的取值范围内，Y 的基本形状保持平坦(均匀)不变。这是独立随机变量定义的性质：无论一个随机变量的输出如何改变，其他随机变量的边缘概率密度函数的形状保持不变(虽然为保证二维概率密度函数的积分为 1，它有一个缩放)。

接下来，我们将定义两个随机变量 X 和 Y 的条件分布，当变量彼此不独立的时候，这是一个很有用的概率测度和相关度指标。

4.5　条件分布

随机变量 X 的条件分布是测度空间$\{\mathcal{R},\mathcal{B}(\mathcal{R})\}$上以一个事件或其他随机变量为条件

的概率测度。这个条件实际上提供了有关试验的信息，并且修改了样本空间。条件分布的概念与第 2 章中讨论的事件概率类似，这里把它扩展到随机变量的累积分布函数和概率密度函数。

定义(条件分布函数) 假设 $P(E)\neq 0$，以事件 E 为条件，随机变量 X 的**累积分布函数**为：

$$F_{X|E}(x|E) \triangleq \frac{P(X\leqslant x,E)}{P(E)} \tag{4-32}$$

一般情况下，没有必要讨论概率为 0 的事件条件下的概率。$F_{X|E}(x|E)$是随机变量 X 的有效的累积分布函数(概率测度)，它具备第 3 章中所描述的性质。注意，它不是条件事件 E 的概率测度。图 4-10 给出了一个条件累积分布函数，其中事件 A 在抽象样本空间 Ω 中，对应$\{X\leqslant x\}$，X 为随机变量。由于我们感兴趣的是随机变量，事件 E 可能代表$\{Y\leqslant y\}$，Y 为另一个随机变量。但也有可能对其他条件感兴趣，例如$\{0\leqslant Y\leqslant y\}$、$\{Y>y\}$等。当事件 $E=(Y\leqslant y)$发生时，条件累积分布函数为：

$$F_{X|E}(x|E) = \frac{P(X\leqslant x,Y\leqslant y)}{P(Y\leqslant y)} = \frac{F_{X,Y}(x,y)}{F_Y(y)} \tag{4-33}$$

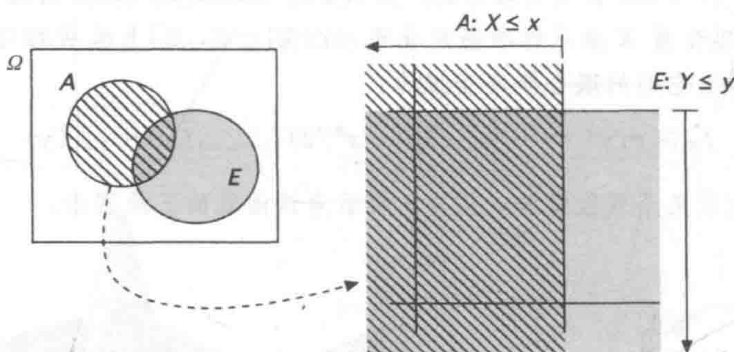


图 4-10 抽象样本空间 Ω 中的条件事件 E 以及 E 到随机变量 Y 的映射

在抽象样本空间中，以 E 为条件使 Ω 缩小为 E，因此只需考虑 A 中包含 E 的一个子集，如图 4-10 中事件的交集所示。这些事件映射至 x-y 平面上，交集 $A\cap E$ 映射到了$\{X\leqslant x\}\cap\{Y\leqslant y\}$的区域。式(4-33)中的联合累积分布函数就定义在这个区域，经过$F_Y(y)$的归一化后，条件累积分布函数变成了一个有效的概率测度。

如果 $F_{X|E}(x|E)$关于 x 绝对连续，利用第 3 章的拉东-尼柯迪姆(Radon-Nikodým)定理，可以得出以下条件概率密度：

$$f_{X|E}(x|E) = \frac{\mathrm{d}}{\mathrm{d}x}F_{X|E}(x|E) \tag{4-34}$$

将式(4-33)代入，对 $E=(Y\leqslant y)$有：

$$f_{X|E}(x|E) = \frac{1}{F_Y(y)}\frac{\partial}{\partial x}F_{X,Y}(x,y) \tag{4-35}$$

需要注意的是，微分没有写成联合概率密度函数 $f_{X,Y}(x,y)$，这是因为我们没有关于条件随机变量 y 求微分。下面将条件分布的定义扩展到关于随机变量 Y 的条件。

我们从定义在一个时间间隔上的概率开始，它是一个具有非零概率的事件，然后令间隔长度趋于 0，对于小 $\varepsilon>0$：

$$F_{X|Y}(x|y\leqslant Y\leqslant y+\varepsilon) = \frac{P(X\leqslant x,y\leqslant Y\leqslant y+\varepsilon)}{P(y\leqslant Y\leqslant y+\varepsilon)} = \frac{\int_{-\infty}^{x}\int_{y}^{y+\varepsilon}f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x}{P(y\leqslant Y\leqslant y+\varepsilon)} \tag{4-36}$$

代入：

$$P(y \leqslant Y \leqslant y+\varepsilon) \approx \varepsilon f_Y(y) \tag{4-37}$$

它是用矩形面积近似 $f_Y(y)$ 在 ε 区间上的积分，可以得到以下近似：

$$F_{X|Y}(x|y \leqslant Y \leqslant y+\varepsilon) \approx \frac{1}{\varepsilon f_Y(y)}\int_{-\infty}^{x}\int_{y}^{y+\varepsilon} f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x \tag{4-38}$$

同样，可以代入：

$$\int_{y}^{y+\varepsilon} f_{X,Y}(x,y)\mathrm{d}y \approx \varepsilon f_{X,Y}(x,y) \tag{4-39}$$

令 $\varepsilon \to 0$，得出：

$$\begin{aligned} F_{X|Y}(x|y) &= \lim_{\varepsilon\to 0} F_{X|Y}(x|y \leqslant Y \leqslant y+\varepsilon) \\ &= \lim_{\varepsilon\to 0}\frac{\int_{-\infty}^{x}\varepsilon f_{X,Y}(x,y)\mathrm{d}y}{\varepsilon f_Y(y)} = \frac{\int_{-\infty}^{x} f_{X,Y}(x,y)\mathrm{d}y}{f_Y(y)} \end{aligned} \tag{4-40}$$

在最后一个等式里，分子、分母消去了 ε。对式(4-40)关于 x 求微分，得出条件概率密度：

$$f_{X|Y}(x|y) = \frac{\partial}{\partial x}F_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{4-41}$$

式(4-41)是在随机变量 $Y=y$ 条件下得出的，而式(4-35)是在随机事件 E(Y 的一个区间)条件下得出的，注意二者的区别。

例 4-5 考虑两个随机变量的均值都为 0，方差都为 1 的二维高斯分布：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right) \tag{4-42}$$

其中 ρ 为互相关系数。Y 的边缘概率密度为：

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\exp(-y^2/2) \tag{4-43}$$

将这两个式子相除，得到条件概率密度函数：

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)+y^2/2\right) \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{1}{2(1-\rho^2)}(x-\rho y)^2\right) \end{aligned} \tag{4-44}$$

它仍然服从高斯分布。与 X 的边缘概率密度函数是标准高斯分布相比，我们发现条件概率密度的均值从 0 变成了 $\mu_{X|Y}=\rho y$，方差从 1 变成了 $\sigma^2_{X|Y}=1-\rho^2$。由于 X、Y 相关，条件中提供了 X 的“信息”。特别是，观察到当 $Y=y$ 时，X 的条件概率的均值为 ρy。图 4-11 给出了 $\rho=0.9$、$y=5$ 时的边缘和条件高斯概率密度函数。在此条件下，由于 Y 与 X 高度相关(有大的正相关系数)，可以得出 X(条件)的输出更可能是在 $5\rho=0.45$ 附近。这是由于方差为 $1-\rho^2=0.19$，与边缘概率密度函数的 1 相比小了很多。另一方面，如果 $\rho=0$，也就是 X 和 Y 是不相关的(进而独立，因为它们是联合高斯的)，那么 X 的条件概率密度函数与它的边缘概率密度函数相同，Y 没有提供 X 的任何信息。◀

在许多应用中，可以利用这种条件信息作出决策或预测，这部分内容将在本书第三部分介绍。本节最后简要介绍信息论中的一个专题(见第 10 章)。

例 4-6 随机变量 X 和 Y 的平均互信息为：

$$I(X;Y) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log(f_{X|Y}(x|y)/f_X(x))\mathrm{d}x\mathrm{d}y \tag{4-45}$$

对数通常取 2 或 e。$I(X;Y)$ 是 X“包含”Y 的信息的测度，反之亦然。观察发现，如果 X 和 Y 不相关，则 $f_{X|Y}(x|y)=f_X(x)$，并且

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log(f_X(x)/f_X(x))\mathrm{d}x\mathrm{d}y = 0 \tag{4-46}$$

当两个随机变量不相关时，它们没有提供关于彼此的任何信息。对于例 4-5 中的联合高斯分布的随机变量，可以直接得出(见习题 4-14)：

$$I(X;Y)=-(1/2)\ln(1-\rho^2) \tag{4-47}$$

图 4-12 画出了该表达式，表明随着相关性的增加(正或负 ρ)互信息显著增加。第 5 章将进一步探讨相关性，讨论相关性如何影响联合分布的形状。◀

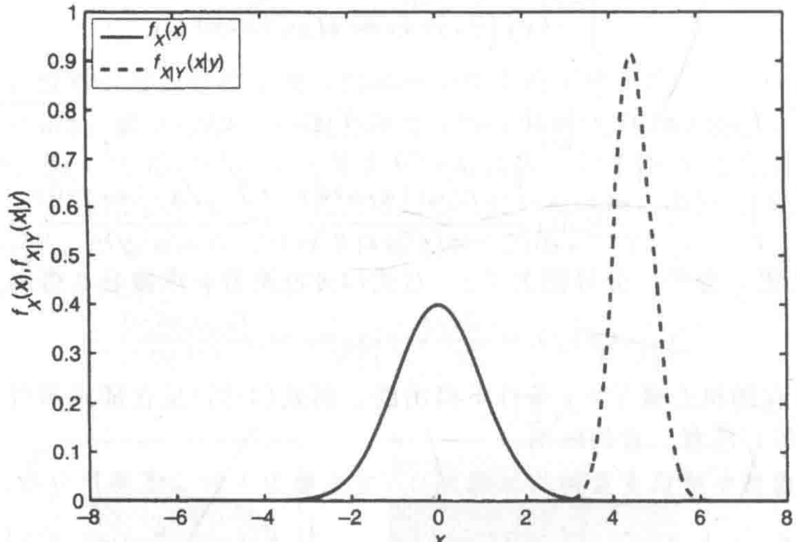


图 4-11　例 4-5 对应的边缘、条件高斯概率密度函数。参数 $\{\mu=0,\sigma=1\}$、$\{\mu_{X|Y}=5,\sigma_{X|Y}=0.4359\}$。条件概率密度函数(点画线)的方差较小

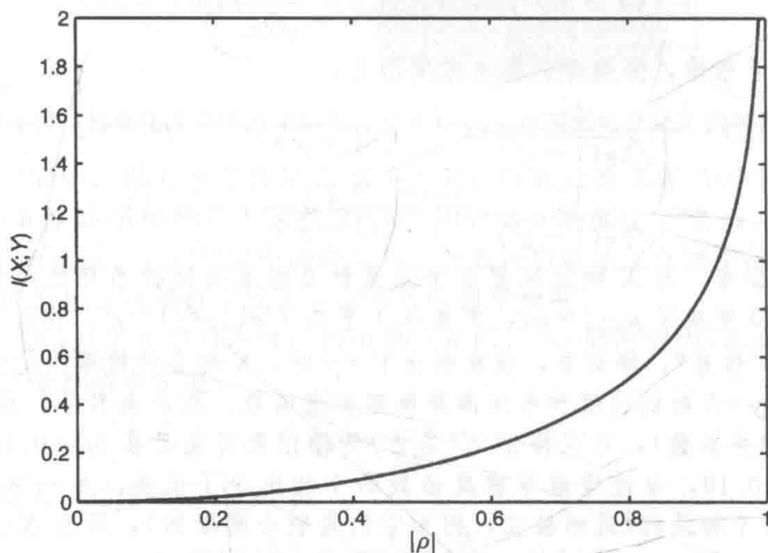


图 4-12　例 4-6 中不同相关系数 ρ 对应的标准高斯随机变量的互信息

4.6 随机向量

可以通过事件空间 $\{\Omega,\mathcal{F}\}$ 的输出到 N 维欧几里得空间 $\mathcal{R}^N=\mathcal{R}\times\cdots\times\mathcal{R}$ 的多维映射(函数)，将随机变量的定义扩展到随机向量。用函数的列向量 $\boldsymbol{X}(\omega)\triangleq[X_1(\omega),\cdots,X_N(\omega)]^{\mathrm{T}}$ 表示多维映射，其中 $X_n(\omega)$ 表示从 $\omega\in\Omega$ 到 $\mathcal{R}^N$ 的第 n 个坐标的映射。因此，向量映射是从事件到 $\mathcal{R}^N$ 中超矩形开集的映射。可简写为：

$$\boldsymbol{X}(\omega)\rightarrow\boldsymbol{X}\triangleq[X_1,\cdots,X_N]^{\mathrm{T}} \tag{4-48}$$

这里 $X\in\mathcal{R}^N$ 表示随机向量。图 4-13 给出了 $N=2$ 时映射的一个例子，对应事件 $A\in\Omega$ 到 $(-\infty,a]\times(-\infty,b]\in\mathcal{R}^2$ 半无限矩形的映射。图中 B 表示$\mathcal{B}(\mathcal{R}^N)$中的超矩形，由笛卡儿积 $\prod_{n=1}^{N}[a_n,b_n](b_n>a_n)$ 确定。尽管$\mathcal{B}(\mathcal{R})$的定义基于$\mathcal{R}$上的开区间，但是它也包括半开区间、闭区间以及离散的点。同理，对 $N>1$ 的随机向量：$\mathcal{B}(\mathcal{R}^N)$包括$\mathcal{R}^N$ 上的全开、半开、闭超矩形以及离散点。在下列两个条件下，映射向量 $\boldsymbol{X}(\omega)$是可测量的：（ⅰ）各部分都是可测量函数（与标量情况相同）（ⅱ）各逆象$\{X_n^{-1}(B)\}$的交集包含在$\mathcal{F}$中：

$$\begin{aligned}\boldsymbol{X}^{-1}(B) &= \{\omega: X_n(\omega)\in[a_n,b_n], n=1,\cdots,N\}\\ &= \bigcap_{n=1}^{N} X_n^{-1}([a_n,b_n]) \triangleq A\in\mathcal{F}\end{aligned} \tag{4-49}$$

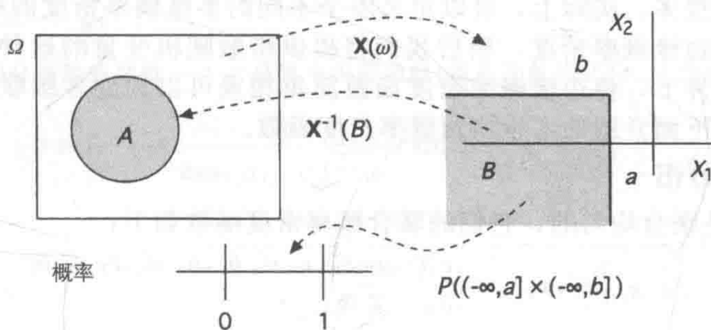


图 4-13　从 Ω 到$\mathcal{R}^2$ 的映射，事件 B 的原象以及到 $[0,1]$ 闭区间的概率映射

第二个条件非常重要：交集定义了 $B\in\mathcal{R}^N$ 代表的超矩形，该矩形必须在抽象样本空间 Ω 中有逆象 $A\in\mathcal{F}$。

例 4-7　考虑抛两个硬币的试验，并定义以下随机变量：$X_1(\omega)=$ 正面 H 的次数，$X_2(\omega)=$ 正面 H 的次数减去反面 T 的次数。很明显，原始样本空间 Ω 里有 $2^2=4$ 个等概率输出，其幂集$\mathcal{P}(\Omega)=\mathcal{F}$有 $2^4=16$ 个元素。随机变量 X_1 输出为 $\{0,1,2\}$，X_2 输出为 $\{-2,0,2\}$，如图 4-14 所示。虽然笛卡儿乘积 $\{0,1,2\}\times\{-2,0,2\}\in\mathcal{R}^2$ 有 9 个点，但只有三个具有非零概率：$\boldsymbol{x}=\{[2,2]^{\mathrm{T}},[1,0]^{\mathrm{T}},[0,-2]^{\mathrm{T}}\}$，见图中实心圆。概率密度函数分别为：

$$f_{X_1}(x_1)=(1/4)\delta(x_1)+(1/2)\delta(x_1-1)+(1/4)\delta(x_1-2) \tag{4-50}$$

$$f_{X_2}(x_2)=(1/2)\delta(x_2)+(1/4)\delta(x_2-2)+(1/4)\delta(x_2+2) \tag{4-51}$$

$$\begin{aligned}f_{X_1,X_2}(x_1,x_2) &= (1/4)\delta(x_1-2,x_2-2)+(1/2)\delta(x_1-1,x_2)\\ &\quad+(1/4)\delta(x_1,x_2+2)\end{aligned} \tag{4-52}$$

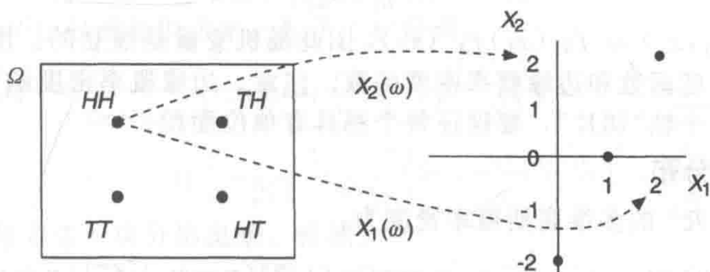


图 4-14　例 4-7 中随机向量 HH 的两个映射，对应到$\mathcal{R}^2$ 中 $\boldsymbol{x}=[2,2]^{\mathrm{T}}$ 输出

其中 $\delta(x_1,x_2)=\delta(x_1)\delta(x_2)$（见附录 B）。通过观察得出边缘概率密度，并且可以通过简单地对联合概率密度函数积分验证它们。需要注意的是，X_1 和 X_2 不是独立的：$f_{X_1,X_2}(x_1,x_2)\neq f_{X_1}(x_1)f_{X_2}(x_2)$。接下来，用 B 表示事件 $\{X_1<2,X_2<0\}$，对应是半无

限矩形，只包括输出$[0,-2]^{\mathrm{T}}$。其逆象是：

$$
\begin{aligned}
\boldsymbol{X}^{-1}(B) &= X_1^{-1}((-\infty,2)) \cap X_2^{-1}((-\infty,0)) \\
&= \{TT,TH,HT\} \cap TT = TT \in \mathcal{F}
\end{aligned} \tag{4-53}
$$

这是Ω空间中的一个特定输出。再考虑另一个事件C：$X_1 \geqslant 1$，其逆象为：

$$
\boldsymbol{X}^{-1}(C) = X_1^{-1}([1,\infty)) = \{TH,HT,HH\} \in \mathcal{F} \tag{4-54}
$$

我们认为它是幂集$\mathcal{P}(\Omega)$的一个事件(不是单个输出)。◀

符号方面，用加粗的大写字母$\boldsymbol{X}$表示随机向量，其输出记做加粗的小写字母$\boldsymbol{x}$。然而，我们也用加粗的大写字母表示(确定)矩阵，如后面将讨论的$\boldsymbol{X}$的自相关矩阵$\boldsymbol{R}_{XX}$。读者应能够从上下文中区分出随机向量$\boldsymbol{X}$和确定矩阵$\boldsymbol{R}_{XX}$。

接下来的几节将介绍一些多维随机变量。事实证明，用途广泛的相关随机变量的多维概率密度函数不是很多。实际上，可以定义多个不同的多维概率密度的版本，使它们都具有相同分布类型的边缘概率密度。稍后我们将提供指数随机变量的这种例子。另一方面，随机变量独立的条件下，将边缘概率密度函数简单相乘可以得到多维联合概率密度函数。其中一个例子就是下面介绍的二维均匀概率密度函数。

4.6.1 二维均匀分布

设$\{X_1, X_2\}$是联合均匀的，它们的联合概率密度函数如下：

$$
\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= \begin{cases} c, & a_1 \leqslant x_1 \leqslant b_1, a_2 \leqslant x_2 \leqslant b_2 \\ 0, & \text{其他} \end{cases} \\
&= cI_{[a_1,b_1]}(x_1)I_{[a_2,b_2]}(x_2)
\end{aligned} \tag{4-55}
$$

由于一个有效的概率密度函数积分值要等于1，所以

$$
\int_{a_1}^{b_1}\int_{a_2}^{b_2} f_{X_1,X_2}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 = c(b_1-a_1)(b_2-a_2) = cL([b_1-a_1]\times[b_2-a_2]) \tag{4-56}
$$

因此常数$c=1/[(b_1-a_1)(b_2-a_2)]$。勒贝格测度$L(\cdot)$给出了由笛卡儿积$[b_1-a_1]\times[b_2-a_2]$定义的矩形区域。从多维概率密度函数得到X_1的边缘概率密度函数为：

$$
f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\mathrm{d}x_2 = \frac{1}{(b_2-a_2)(b_1-a_1)}\int_{-\infty}^{\infty} I_{[a_1,b_1]}(x_1)I_{[a_2,b_2]}(x_2)\mathrm{d}x_2 \tag{4-57}
$$

X_2的指示函数确定了积分范围，因此

$$
f_{X_1}(x_1) = \frac{1}{b_1-a_1}I_{[a_1,b_1]}(x_1) \tag{4-58}
$$

同理对X_2：

$$
f_{X_2}(x_2) = \frac{1}{b_2-a_2}I_{[a_2,b_2]}(x_2) \tag{4-59}
$$

由于$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$，因此随机变量是独立的。图4-15给出了二维均匀分布的概率密度函数和边缘概率密度函数。注意，边缘概率密度函数不是简单地沿着二维概率密度的两个轴“切片”，要保证每个都具有单位面积。

4.6.2 多维高斯分布

随机向量$X\in\mathcal{R}^N$的多维高斯概率密度为：

$$
f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}_{\boldsymbol{XX}}|^{1/2}}\exp(-(1/2)(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}})^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{XX}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{X}})) \tag{4-60}
$$

其中$\mu \triangleq [\mu_1,\cdots,\mu_N]^{\mathrm{T}}$，$\mu_n$为$X_n$的均值，$\boldsymbol{C}_{\boldsymbol{XX}}=\mathcal{E}[(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})^{\mathrm{T}}]$是一个协方差矩阵，$\mathcal{E}[\cdot]$是期望(详见第5章)，$\boldsymbol{C}_{\boldsymbol{XX}}^{-1}$是它的逆矩阵(假设它是奇异的)，$|\boldsymbol{C}_{\boldsymbol{XX}}|$是行列式。图4-8给出了$N=2$(随机变量$X$和$Y$)的概率密度函数。相应的二维高斯概率密度函数经常写成如下标量形式：

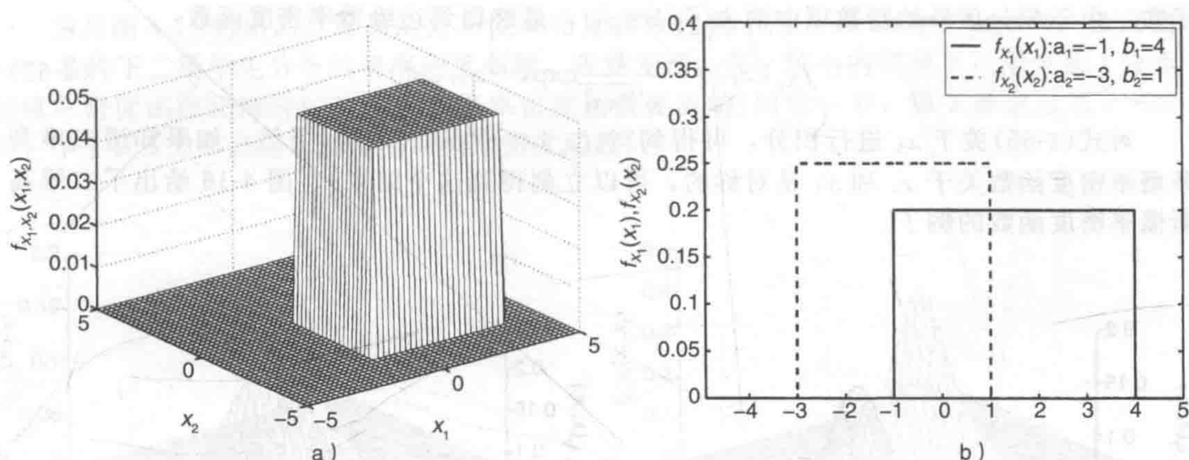


图 4-15　均匀分布随机变量。a) 二维概率密度函数(随机变量独立)；b) 边缘概率密度函数

$$f_{X_1,X_2}(x_1,x_2)=\frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_{X_1})^2}{\sigma_{X_1}^2}-\frac{2\rho(x_1-\mu_{X_1})(x_2-\mu_{X_2})}{\sigma_{X_1}\sigma_{X_2}}+\frac{(x_2-\mu_{X_2})^2}{\sigma_{X_2}^2}\right)\right) \tag{4-61}$$

其中相关系数为：

$$\rho\triangleq\mathcal{E}[(X_1-\mu_{X_1})(X_2-\mu_{X_2})]/\sigma_{X_1}\sigma_{X_2} \tag{4-62}$$

这些参数与式(4-60)中 $\boldsymbol{C}_{XX}$ 的关系是：

$$\boldsymbol{C}_{XX}=\begin{bmatrix}\sigma_{X_1}^2 & \rho\sigma_{X_1}\sigma_{X_2}\\ \rho\sigma_{X_1}\sigma_{X_2} & \sigma_{X2}^2\end{bmatrix} \tag{4-63}$$

边缘概率密度通过对 $f_{X_1X_2}(x_1,x_2)$ 积分得到，与 4.5 节对二维均匀概率密度所做的一样。为了使大家更容易理解推导过程，考虑 $\mu_{X_1}=\mu_{X_2}=0$ 的情况。为符号描述方便，令 $\sigma_1^2\equiv\sigma_{X_1}^2$、$\sigma_2^2\equiv\sigma_{X_2}^2$。则

$$\begin{aligned}f_{X_1}(x_1)&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2}-\frac{2\rho x_1x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right)\right)\mathrm{d}x_2\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{x_1^2}{2\sigma_1^2(1-\rho^2)}\right)\\&\quad\cdot\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2(1-\rho^2)}\left(-\frac{2\rho x_1x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right)\right)\mathrm{d}x_2\end{aligned} \tag{4-64}$$

其中不包含 x_2 的项已经被提取出来。配平方后得到：

$$\begin{aligned}f_{X_1}(x_1)&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{x_1^2}{2\sigma_1^2(1-\rho^2)}\right)\\&\quad\cdot\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2(1-p^2)}\left[\left(\frac{x_2-\rho x_1\sigma_2/\sigma_1}{\sigma_2}\right)^2-\frac{\rho^2x_1^2}{\sigma_1^2}\right]\right)\mathrm{d}x_2\end{aligned} \tag{4-65}$$

将指数里面最右边一项分解出来，得到：

$$\begin{aligned}f_{X_1}(x_1)&=\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{x_1^2-\rho^2x_1^2}{2\sigma_1^2(1-\rho^2)}\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}\\&\quad\cdot\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_2-\rho x_1\sigma_2/\sigma_1}{\sigma_2}\right)^2\right)\mathrm{d}x_2\end{aligned} \tag{4-66}$$

积分值为 1，因为被积函数是均值为 $\rho x_1\sigma_2/\sigma_1$、方差为 $(1-\rho^2)\sigma_2^2$ 的高斯分布概率密度

函数。由于积分项外的指数项中消去了$(1-\rho^2)$，最终得到边缘概率密度函数：

$$f_{X_1}(x_1)=\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp(-x_1^2/2\sigma_1^2) \tag{4-67}$$

对式(4-65)关于x_1进行积分，可得到对$f_{X_2}(x_2)$类似的结果(当然，如果知道二维高斯概率密度函数关于x_1和x_2是对称的，可以立刻得到这个结果)。图4-16给出了二维高斯概率密度函数的例子。

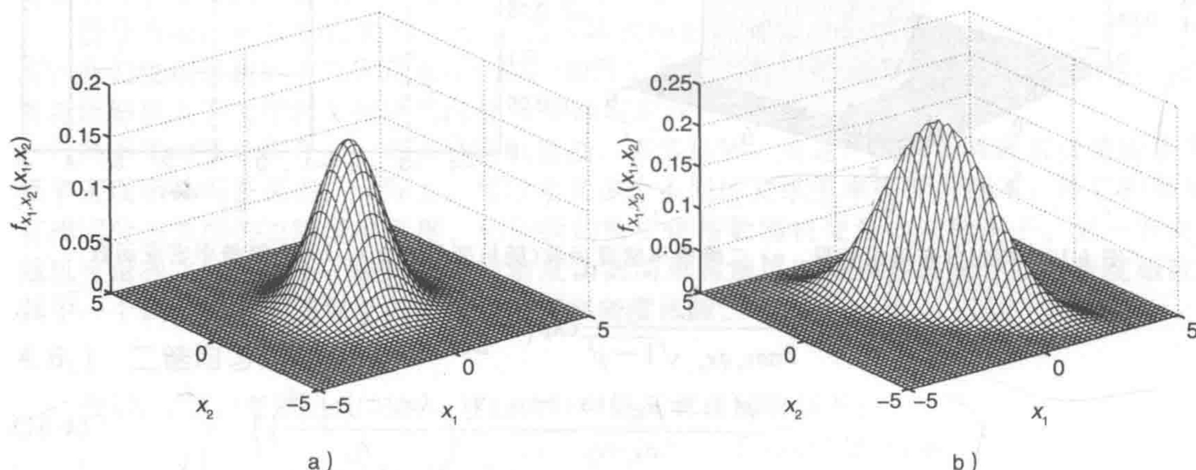


图4-16 $\mu_{X_1}=\mu_{X_2}=0$、$\sigma_{X_1}^2=1$条件下二维高斯概率密度函数。a) $\sigma_{X_2}^2=1$、$\rho=0$(随机变量独立)；b) $\sigma_{X_2}^2=2$、$\rho=-0.8$

当$N>1$时，一般情况下任何低阶联合高斯分布概率密度都可以通过只保留感兴趣的随机变量的对应参数获得。例如，如果原来的联合概率密度函数有三个变量$\{X_1,X_2,X_3\}$，而我们感兴趣的是$\{X_2,X_3\}$的联合概率密度函数，只需要进行如下改变：

$$\boldsymbol{X}=[X_1,X_2,X_3]^{\mathrm{T}}\rightarrow[X_2,X_3]^{\mathrm{T}} \tag{4-68}$$

$$\boldsymbol{\mu_X}=[\mu_{X_1},\mu_{X_2},\mu_{X_3}]^{\mathrm{T}}\rightarrow[\mu_{X_2},\mu_{X_3}]^{\mathrm{T}} \tag{4-69}$$

$$\boldsymbol{C_{XX}}=\begin{bmatrix}C_{11}&C_{12}&C_{13}\\C_{21}&C_{22}&C_{23}\\C_{31}&C_{32}&C_{33}\end{bmatrix}\rightarrow\begin{bmatrix}C_{22}&C_{23}\\C_{32}&C_{33}\end{bmatrix} \tag{4-70}$$

$(2\pi)^{N/2}$的指数项里，N的次数由3变成2，以保证提供正确的归一化使得低阶随机向量的联合概率密度函数的体积为1。

4.6.3 多维学生氏 t 分布

具有N个分量的学生氏t分布的多维概率密度函数为：

$$f_{\boldsymbol{X}}(x)=\frac{\Gamma((r+N)/2)}{\Gamma(r/2)(r\pi)^{N/2}\,|\,\widetilde{\boldsymbol{C}}_{\boldsymbol{XX}}\,|^{1/2}}(1+(\boldsymbol{x}-\boldsymbol{\mu_X})^{\mathrm{T}}\,\widetilde{\boldsymbol{C}}_{\boldsymbol{XX}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_X})/r)^{-(r+N)/2} \tag{4-71}$$

其中$r>0$是自由度。$\widetilde{\boldsymbol{C}}_{\boldsymbol{XX}}$是协方差矩阵，其主对角线的值为1，其他元素与多维高斯概率密度函数的$\boldsymbol{C_{XX}}$的值相等。当且仅当所有随机变量有单位方差时，$\widetilde{\boldsymbol{C}}_{\boldsymbol{XX}}$和$\boldsymbol{C_{XX}}$相等。$N=2$时，二维概率密度函数为：

$$\begin{aligned}f_{X_1,X_2}(x_1,x_2)=&\frac{\Gamma((r+2)/2)}{\Gamma(r/2)k\pi\sqrt{1-\rho^2}}(1+[1/r(1-\rho^2)][(x_1-\mu_{X_1})^2\\&-2\rho(x_1-\mu_{X_1})(x_2-\mu_{X_2})+(x_2-\mu_{X_2})^2])^{-(r+2)/2}\end{aligned} \tag{4-72}$$

这种扩展形式利用了：

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{XX}}=\begin{bmatrix}1&\rho\\\rho&1\end{bmatrix} \tag{4-73}$$

为与图 4-16 给出的二维高斯分布概率密度函数比较，图 4-17 给出了 $\boldsymbol{\mu}_X=0$、自由度 $r=1$ 条件下二维学生分布的概率密度函数。观察发现，在 r 较小的情况下，学生氏 t 分布的概率密度函数比对应的高斯分布概率密度函数要陡峭（回想一下，第 3 章学过当 $r\to\infty$ 时，学生氏 t 分布趋近于标准高斯概率密度函数）。

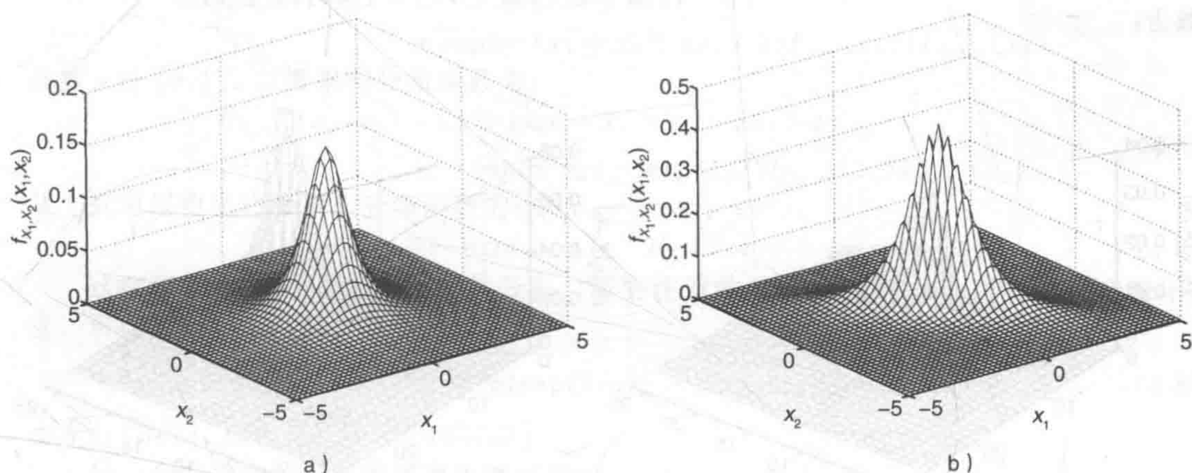


图 4-17 当 $\mu_{X_1}=\mu_{X_2}=0$，$r=1$ 时的二维学生氏 t 分布概率密度函数。a) $\rho=0$（随机变量独立）；b) $\rho=-0.8$

4.6.4 多项分布

多项分布是有 M 个不同输出的二项分布的扩展。有 M 个随机变量 $\{X_1,\cdots,X_M\}$，不同变量出现的概率分别是 $\{p_1,\cdots,p_M\}\in[0,1]$，满足 $\sum_{m=1}^{M}p_m=1$。对 N 次试验，当 $\sum_{m=1}^{M}X_m=N$ 时，联合概率质量函数为：

$$
\begin{aligned}
p_{X_1,\cdots,X_M}[x_1,\cdots,x_M]&=N!\prod_{m=1}^{M}\frac{p_m^{x_m}}{x_m!}I_{\{N\}}\left[\sum_{m=1}^{M}x_m\right]\\
&=\binom{N}{x_1,\cdots,x_m}\prod_{m=1}^{M}p_m^{x_m}I_{\{N\}}\left[\sum_{m=1}^{M}x_m\right]
\end{aligned}
\tag{4-74}
$$

其中，

$$
\binom{N}{x_1,\cdots,x_M}\triangleq\frac{N!}{\prod_{m=1}^{M}x_m!}=\frac{\Gamma(N+1)}{\prod_{m=1}^{M}\Gamma(x_m+1)}
\tag{4-75}
$$

是多项式系数，$\Gamma(\cdot)$ 为伽马函数。第 m 个分量的均值 $\mu_{X_m}=Np_m$，方差 $\sigma_{X_m}^2=Np_mq_m$，其中 $q_m\triangleq1-p_m$。随机变量总是负相关：

$$
\rho_{X_mX_n}=-\sqrt{p_mp_n/q_mq_n}
\tag{4-76}
$$

多维累积分布函数是：

$$
F_{x_1},\cdots,x_M(x_1,\cdots,x_M)=N!\sum_{n_1=0}^{x_1}\cdots\sum_{n_{M-1}=0}^{x_{M-1}}\frac{1}{n_1!\cdots\left(N-\sum_{m=1}^{M-1}n_m\right)!}\prod_{m=1}^{M}p_m^{n_m}
\tag{4-77}
$$

只需要计算 $M-1$ 个和，因为 x_M 由式(4-74)中的指示函数确定，也就是 $x_M=1-\sum_{m=1}^{M-1}x_m$。图 4-18 给出了 $M=3$（仅示出 x_1 和 x_2）的概率质量函数的例子。

4.6.5 多维超几何分布

多维超几何分布是将两种对象类型的超几何概率质量函数扩展到 $L\in\mathcal{N}$ 个对象类型。总共有 $N\in\mathcal{N}$ 个对象不变，类型 1 有 M_1 个对象，类型 2 有 M_2 个对象，依此类推，类型 L 有 M_L 个

对象，则 $\sum_{m=1}^{L} M_m = N_L$。(对于超几何分布，M 个对象是一种类型，$N-M$ 个对象是另一种类型。)多维超几何分布模型为抽取后不放回，即 $n \leqslant N$。而多项分布假设抽取后放回，即 N 保持不变。因此，它们的关系与超几何和二项分布随机变量类似。n 个样本的多维概率质量函数为：

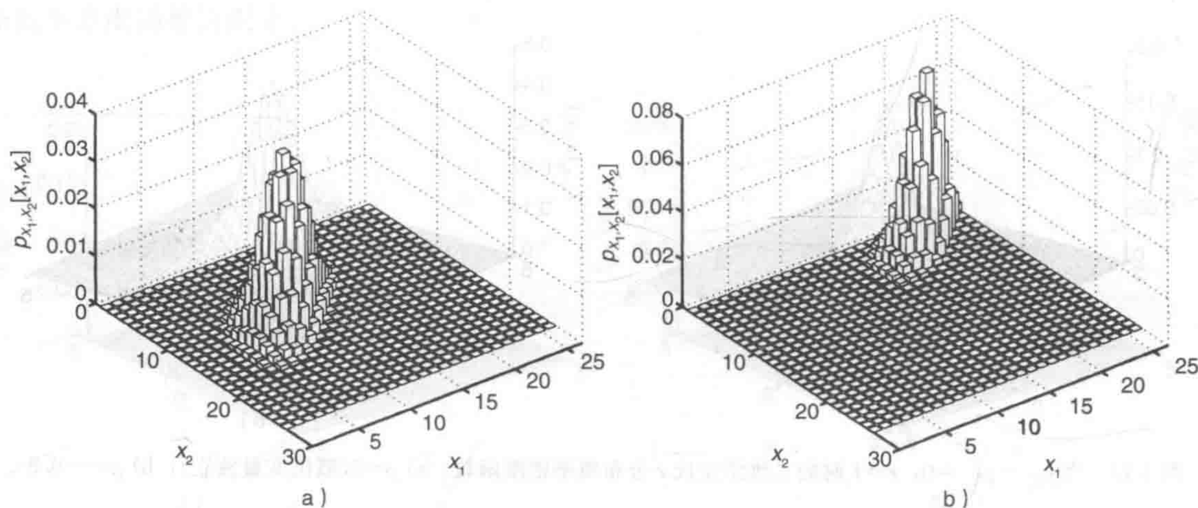


图 4-18 $M=3$ 的多项分布概率质量函数(只给出了 x_1、x_2)。a) $N=25$、$\{p_1,p_2,p_3\}=\{0.4,0.4,0.2\}$；b) $N=25$、$\{p_1,p_2,p_3\}=\{0.8,0.1,0.1\}$

$$p_{X_1,\cdots,X_L}[x_1,\cdots,x_L]=\frac{\prod_{m=1}^{L}\binom{M_m}{x_m}}{\binom{N}{n}}I_{(n)}\left[\sum_{m=1}^{L}x_m\right]I_{(0,\cdots,N)}[n]\prod_{m=1}^{L}I_{(0,\cdots,M_m)}[x_m] \tag{4-78}$$

其中指示函数指定了所有值的范围。第 m 个随机变量的均值 $\mu_{X_m}=nM_m/N$，方差 $\sigma_{X_m}^2=nM_m(N-M_m)(N-n)/N^2(N-1)$。与多项分布一样，随机变量总是呈负相关：

$$\rho_{X_kX_l}=-\sqrt{M_kM_l/(N-M_k)(N-M_l)} \tag{4-79}$$

图 4-19 给出了 $L=3$(仅示出 x_1 和 x_2)的概率质量函数。需要注意的是，输出 x_n 不能超过 M_n，分布在每个方向上比较窄，这是因为 n 和 N 很接近(见前面的均值和方差表达式)。

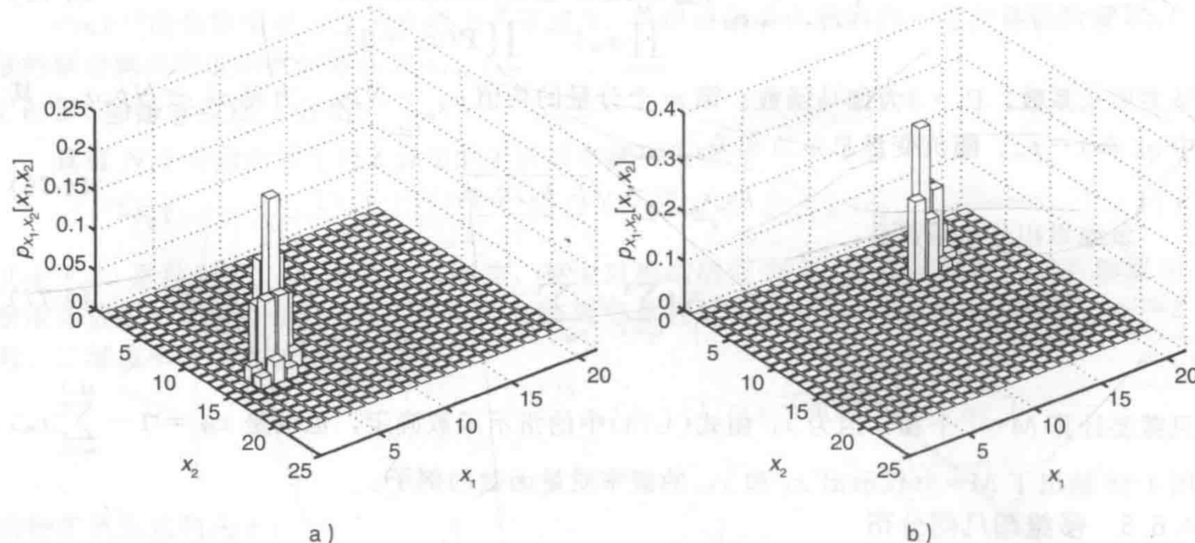


图 4-19 $L=3$ 的多维概率质量函数(只给出了 x_1、x_2)。a) $N=25, n=20, \{M_1,M_2,M_3\}=\{5,15,5\}$；b) $N=25, n=20, \{M_1,M_2,M_3\}=\{18,5,2\}$

4.6.6 二维指数分布

正如前面提到的，可以定义随机向量的不同多维分布，它是单变量情况下的一个扩展。例如，下面给出了一个可能的二维指数分布的概率密度函数(Gumbel，1960)：

$$f_{X_1,X_2}(x_1,x_2)=[(1+\alpha x_1)(1+\alpha x_2)-\alpha] \cdot \exp(-(x_1+x_2+\alpha x_1 x_2))I_{[0,\infty)}(x_1)I_{[0,\infty)}(x_2) \tag{4-80}$$

参数 $\alpha\in[0,1]$。二维累积分布函数为：

$$F_{X_1,X_2}(x_1,x_2)=[1+\exp(-x_1-x_2-\alpha x_1 x_2) - \exp(-x_1)-\exp(x_2)]I_{[0,\infty)}(x_1)I_{[0,\infty)}(x_2) \tag{4-81}$$

从上式可以得出边缘累积分布函数为(令 $x_1\to\infty$、$x_2\to\infty$)：

$$F_{X_1}(x_1)=1-\exp(-x_1),\quad F_{X_2}(x_2)=1-\exp(-x_2) \tag{4-82}$$

这些都是参数 $\lambda=1$ 的标准指数分布。需要注意的是，除非 $\alpha=0$，否则随机变量不独立。可以证明相关系数为：

$$\rho=-1-(1/\alpha)\exp(1/\alpha)\int_1^\infty(1/v)\exp(-v/\alpha)\mathrm{d}v \tag{4-83}$$

由 $\alpha\in[0,1]$，得出 $\rho\in[-0.4037,0]$。

另一个二维指数分布的概率密度函数为：

$$f_{X_1,X_2}(x_1,x_2)=[1+\alpha(2\exp(-x_1)-1)(2\exp(-x_2)-1)] \cdot \exp(-x_1-x_2)I_{[0,\infty)}(x_1)I_{[0,\infty)}(x_2) \tag{4-84}$$

累积分布函数为：

$$F_{X_1,X_2}(x_1,x_2)=(1-\exp(-x_1))(1-\exp(-x_2)) \cdot [1+\alpha\exp(-x_1-x_2)]I_{[0,\infty)}(x_1)I_{[0,\infty)}(x_2) \tag{4-85}$$

其中 $\alpha\in[-1,1]$。与式(4-82)类似，边缘概率密度函数是 $\lambda=1$ 的标准指数分布。相关系数为(见习题 5-35)：

$$\rho=\alpha/4 \tag{4-86}$$

$\rho\in[-1/4,1/4]$。当且仅当 $\alpha=0$ 时随机变量独立。有趣的是，虽然式(4-80)和式(4-84)具有相同的边缘概率密度函数，但是相关系数有很大的不同。图 4-20 给出了 $\alpha=1/2$ 时的二维概率密度函数。有可能存在二维指数概率密度函数的其他版本(见习题 4-11)，例如将另一簇的单随机变量扩展为多维分布。

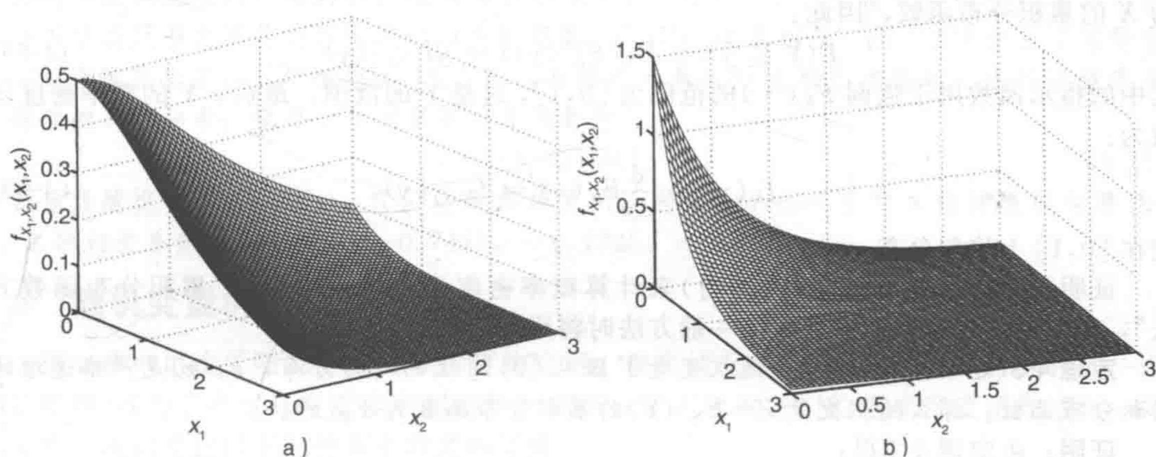


图 4-20 $\alpha=1/2$ 时的二维指数概率密度函数。a) 式(4-80)中的分布；b) 式(4-84)中的分布

4.7 产生相关随机变量

可以利用 $\prod_{n=1}^{N} f_{X_n}(x_n)$，简单地生成 N 个独立边缘概率密度函数的多维概率密度函数

$f_{\boldsymbol{X}}(\boldsymbol{x})$。从前面的例子可以看出，即使边缘概率密度函数服从不同的分布，同样也可以通过相乘得到联合分布。很遗憾，没有很多广泛应用的相关随机变量的多维分布，就如 ρ 明确时二维高斯和学生氏 t 分布的情况。以二维情况($N=2$)为例，可以通过下面的方法，引入其他随机量的相关性。该方法不产生所需联合概率密度函数的闭合表达式，而是计算机模拟样本的一种方法：

- 用 $\boldsymbol{X}=[X_1,X_2]^{\mathrm{T}}$ 表示具有二维标准高斯分布概率密度函数的随机向量，其相关系数 $\rho\neq 0$。注意，对于二维标准高斯分布的概率密度函数，均值为 0，$\boldsymbol{C_{XX}}=\boldsymbol{\tilde{C}_{XX}}$。式(4-73)定义的矩阵与二维学生氏 t 分布相关；不同于 $\boldsymbol{C_{XX}}$，它的主对角线的元素值为 1。随机变量 $\{X_1,X_2\}$ 的边缘分布均为标准高斯分布。
- 对每个随机变量做概率变换 $Y_n=F_X(X_n)$，$n=1$，2，其中 $F_X(\cdot)$为标准高斯随机变量的累积分布函数。$\{Y_n\}$为 $[0,1]$ 上的均匀分布(见定理 4-2)，它们是相关的，因为初始的$\{X_n\}$是相关的。
- 既然随机变量服从均匀分布，并且都可以通过相应累积分布函数的逆(见定理 4-3 和第 1 章中的讨论)转化成一个新的随机变量。因此，得到 $Z_n=F_Z^{-1}(Y_n)$，$n=1$，2，其中 $F_Z^{-1}(\cdot)$是所需随机变量的逆累积分布函数(icdf)。到相关$\{Z_n\}$的转换不需要是一样的。

总结一下：$\{X_1,X_2\}$ 服从相关系数为 ρ 的标准二维高斯分布，通过下式得到变换后的随机变量 $\{Z_1,Z_2\}$：

$$Z_n=F_{Z_n}^{-1}(F_X(X_n)),\quad n=1,2 \tag{4-87}$$

其中 $F_{Z_n}(\cdot)$里的下标 n 用于强调到每个 Z_n 的映射可以不同。虽然该方法能够产生任意二维概率密度函数的相关随机变量，Z_1 和 Z_2 的相关性与相关系数为 ρ 的二维高斯变量 $\{X_1,X_2\}$ 的线性相关性不同(见第 5 章线性关系的讨论)。

定理 4-2(概率变换) 设连续随机变量 X 的累积分布函数 $F_X(X)$是严格递增的，那么随机变量 $Y=F_X(X)$服从 $[0,1]$ 上的均匀分布。

证明：注意，$F_X(X)$的变量是随机变量 X(而不是输出 x)。因此，$F_X(X)$是一个映射，可得：

$$P(Y\leqslant y)=p(F_X(X)\leqslant y)=P(X\leqslant F_X^{-1}(y)) \tag{4-88}$$

其中逆象 $F_X^{-1}(y)$是非常明确的，因为 $F_X(\cdot)$是严格递增(一一对应)的。最后一个表达式为 X 的累积分布函数，因此：

$$P(Y\leqslant y)=F_X(F_X^{-1}(y))=yI_{[0,1]}(y) \tag{4-89}$$

其中的指示函数用于强调 $F_X(\cdot)$的范围为$[0,1]$，这是 Y 的范围。最后，Y 的概率密度函数为：

$$f_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}F_Y(y)=I_{[0,1]}(y) \tag{4-90}$$

它在 $[0,1]$ 上均匀分布。

证明过程中使用的由变换(映射)来计算概率密度函数的方法称为“累积分布函数方法”，本章后面讨论随机变量变换一般方法时将用到该知识点。

定理 4-3(逆概率变换) 令随机变量 Y 服从 $[0,1]$ 上的均匀分布，$F_z(z)$是严格递增的累积分布函数。那么随机变量 $Z=F_X^{-1}(Y)$的累积分布函数为 $F_Z(z)$。

证明：由定理 4-2 得：

$$P(Z\leqslant z)=p(F_Z^{-1}(Y)\leqslant z)=P(Y\leqslant F_Z(z))=F_Y(F_Z(z)) \tag{4-91}$$

由于 $F_Y(y)=yI_{[0,1]}(y)$是一个关于自变量的线性函数，表达式简化为：

$$P(Z\leqslant z)=F_Z(z)I_{[0,1]}(F_Z(z))=F_Z(z) \tag{4-92}$$

由定义知道 $F_Z(z)$的范围是$[0,1]$，因此指示函数可以去掉。

前面定理中，要求累积分布函数为严格递增的，得到了相对简单的变换结果。对一般

情况下的累积分布函数的变换结果，需要做更多的工作，4.8 节将研究这部分内容。

例 4-8 图 4-21a 给出了二维高斯分布的 100 000 个样本的直方图，其均值向量 $\boldsymbol{\mu}_X=0$，协方差矩阵为

$$C_{XX}=\begin{bmatrix}1 & -0.75\\ -0.75 & 1\end{bmatrix}\tag{4-93}$$

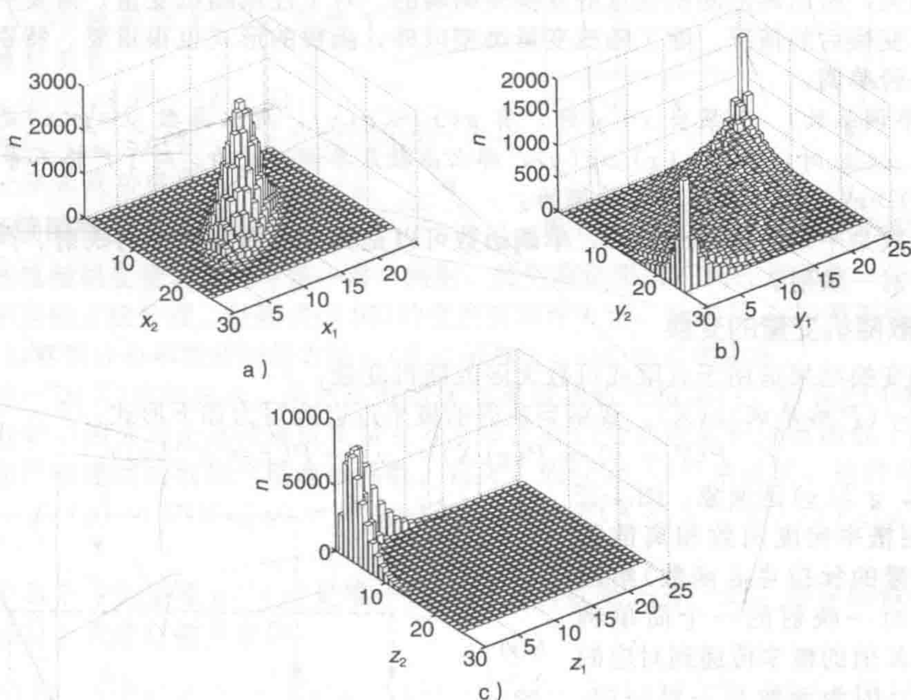


图 4-21　二维随机变量的直方图。a) $\rho=-0.75$ 的二维高斯分布；b) 利用概率变换得到二维均匀分布；c) 利用逆概率变换得到二维指数分布

相关系数 $\rho=-0.75$。利用 MATLAB 中的 mvnrand(它是 randn 的多维版本)函数生成数据。负相关系数使得分布集中在 $x_2=-x_1$ 轴线上。图 4-21b 给出了 $\{X_1,X_2\}$ 分别使用上面的概率变换技术时，得到的二维均匀分布结果。(注意，由于变量是相关的，因此对应的分布在其范围内不是均匀平坦的。)随机变量 $\{Y_1,Y_2\}$ 也负相关。图 4-21c 给出了对每个均匀分布随机变量，利用参数 $\lambda=0.5$ 的指数随机变量的逆累积分布函数，进行逆概率变换得到的二维分布。概率 p 的逆累积分布函数为

$$z=-(1/\lambda)\ln(1-p)\tag{4-94}$$

得到指数随机变量 $\{Z_1,Z_2\}$，它们也是负相关的。利用 corrcoef 由样本估计随机向量 $\boldsymbol{X}$、$\boldsymbol{Y}$、$\boldsymbol{Z}$ 的相关系数，分别为 $\hat{\rho}=-0.7471$，-0.7325，-0.5162。◀

4.8　随机变量的变换

第 3 章中讨论了随机变量很多重要的应用。很多情况下，由于一些物理机制(如一个通信信道)或为了产生具有更多有用特性的随机变量，会对它们的概率密度函数进行一定的变换。我们考虑以下三种基本形式的变换：

- 一个随机变量是另一个随机变量的函数 $Y=g(X)$，其中映射 $g(\cdot)$ 已知且确定。
- 利用不同函数 $\{g_n(\cdot)\}$ 将几个随机变量映射成另一个随机变量集。
- 代表随机过程的一组按时间索引的随机变量集经过滤波系统后，生成了修正的随机过程。

通常情况下，我们感兴趣的是随机变量的概率密度函数在这些变换后是如何改变的。第 8 章中研究系统作用于随机过程时将讨论第三种变换，还将讨论滤波对相关性的影响。

本节主要讨论第一种形式，即单个随机变量的变换：

$$Y = g(X) \tag{4-95}$$

其中 $g(\cdot)$是一个无记忆的映射。至于第二种形式，本章稍后将讨论多个随机变量的特殊映射。

方便起见，分别考虑离散、连续随机变量。由于离散随机变量的概率质量函数与事件概率直接相关，所以离散随机变量的变换是明确的。对于连续随机变量，需要利用微积分知识来计算变换后的情况。除了随机变量类型以外，函数的形式也很重要。特别是，需要知道函数是否单调。

定义（单调函数） 如果当 $x \leqslant y$ 时，有 $g(x) \leqslant g(y)$，那么函数 $y = g(x)$是单调递增的。如果当 $x \leqslant y$ 时，都有 $g(x) \geqslant g(y)$，那么函数是单调递减的。对于严格不等式 $g(x) < g(y)$和 $g(x) > g(y)$，函数是严格单调的。

如果函数是平坦的(微分是零)，单调函数可以是其范围内多到一的映射。严格单调函数也称为一对一映射。

4.8.1 离散随机变量的变换

下面的变换结果适用于有限或可数无限的随机变量：

- 一对一(严格单调)$g(X)$。观察到映射的概率可以改写为如下形式：

$$P(Y = y) = P(g(X) = y) = P(x = g^{-1}(y)) \tag{4-96}$$

式中，$g^{-1}(y)$是逆象。图 4-22(使用概率密度函数和离散随机变量的狄拉克 δ 函数)给出了一对一映射的一个简单例子。X 值的概率传递到对应的 Y 值。因为函数是一对一的，所以 Y 和 X 的输出数目要相同并且相对概率不变。这种类型的映射本质上是对不同输出的概率重置。

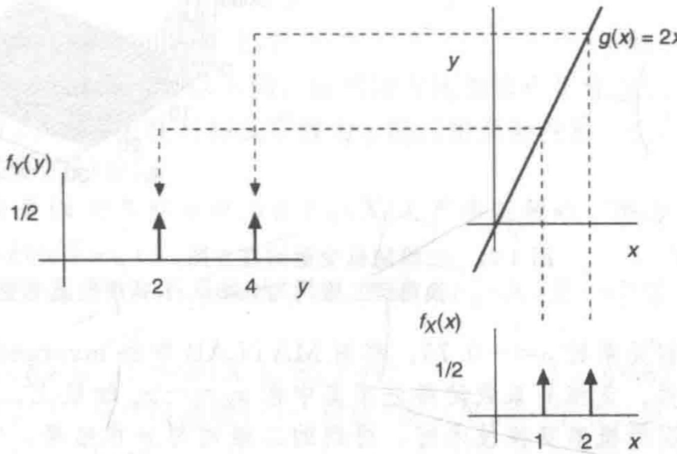


图 4-22 离散随机变量一对一变换的概率映射

一个函数通过定义可以将其自变量映射到单个点，但是这对逆象通常不成立。对一个具有平坦区域的单调函数而言，它是多对一的映射。

- 单调 $g(X)$。该函数沿 x 轴在平坦区域内被分成多段，并且是一对一区域。像上面说的那样，严格单调区域分别处理，将每个 x 的概率映射到对应的 y 值。在恒定值区域，映射到 y 的所有 x 值的累积概率质量即是特定 y 的概率。即使是非单调函数，这个结果也是正确的，但我们必须仔细考虑映射到同一个 y 的所有 x 区域，下面将介绍这一点。
- 任意 $g(X)$。假设 N 个不同 x 值映射到同一个 y 值。由于 X 是离散的并且 x 值不同(互斥)，因此总是可以将概率写成 X 值集合的形式：

$$P(Y = y) = P\Big(\bigcup_{n=1}^{N} \{X = x_n\}\Big) = \sum_{n=1}^{N} P(X = x_n)$$
$$= \sum_{n-1}^{N} P(X = g_n^{-1}(y))$$

其中 $g_n^{-1}(y)$的下标表示映射到同一 y 值的不同 x 值。N 不需要是有限的：结果可以包括映射到相同的 y 值的可数的无穷多个 x 值。我们只需要跟踪到所有的 x 值，

从而对应的累积概率质量被赋予正确的y值。图 4-23 给出了一个有限随机变量的非单调的例子。

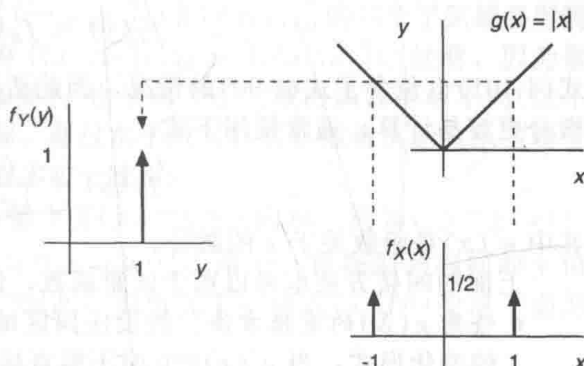


图 4-23　离散变量多对一变换的概率映射

例 4-9　X 是一个参数为 p 的几何分布随机变量，假设映射 $g(X)=u(X-3)$（移位的单位阶跃函数）。因此，$X=\{0,1,2\}$ 映射到 $Y=0$，$X\geqslant 3$ 映射到 $Y=1$。显然 Y 是一个伯努利随机变量：

$$P(Y=0)=p(p^2-3p+3),$$
$$P(Y=1)=1-p(p^2-3p+3) \quad (4\text{-}97)$$

其中最后一项是成功概率。◀

4.8.2　连续随机变量的变换

对于连续随机变量，首先考虑一对一映射，然后将结果推广到任意函数。导数为零的区域必须用不同的方法处理。处理式(4-95)的变换有四种方法，分别是：(ⅰ)概率密度函数或直接方法；(ⅱ)累积分布函数或间接方法；(ⅲ)δ 函数法；(ⅳ)指示函数法。

- 一对一 $g(X)$的间接法。虽然这种情况与离散随机变量类似，但我们不能从 $P(Y=y)$开始，因为对于连续随机变量它等于零。我们会利用累积分布函数 $F_Y(y)$，分别考虑严格递增函数和严格递减函数。首先，假设 $g(X)$严格递增，这样可以写成：

$$F_Y(y)=P(Y\leqslant y)=P(g(X)\leqslant y)=P(X\leqslant g^{-1}(y))=F_X(g^{-1}(y)) \quad (4\text{-}98)$$

其中每个 y 的逆象 $g^{-1}(y)$是唯一的，因为函数是一对一的。相应的概率密度函数通过对上式进行微分获得：

$$f_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}F_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}F_X(g^{-1}(y))=\frac{\mathrm{d}}{\mathrm{d}x}F_X(g^{-1}(y))\frac{\mathrm{d}x}{\mathrm{d}y}=f_X(g^{-1}(y))\frac{\mathrm{d}x}{\mathrm{d}y} \quad (4\text{-}99)$$

其中利用了链式法则并且代入了 $x=g^{-1}(y)$。

对于严格递减函数，式(4-98)的倒数第二个式子变为：

$$F_Y(y)=P(X\geqslant g^{-1}(y))=1-F_X(g^{-1}(y)) \quad (4\text{-}100)$$

不等号方向改变。这一点可以从图 4-24 所示的均匀分布的例子看出。

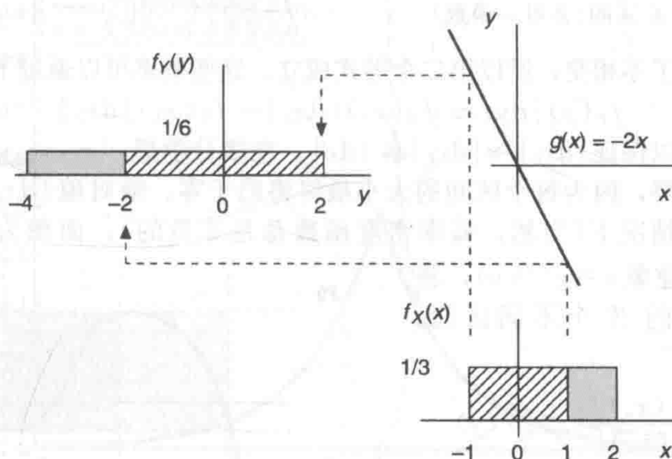


图 4-24　严格递减函数对应的连续随机变量的映射。从阴影区域可以看出 $P(Y\leqslant -2)=P(X\geqslant 1)$，这是因为映射将概率密度函数进行了翻转

对式(4-100)进行微分，得到：

$$f_Y(y)=-f_X(g^{-1}(y))\frac{\mathrm{d}x}{\mathrm{d}y} \quad (4\text{-}101)$$

除了符号相反外，结果与式(4-99)相同。然而，由于微分 $\mathrm{d}x/\mathrm{d}y$ 在严格递减函数条件下为

负，式(4-101)可写成：

$$f_Y(y)=f_X(g^{-1}(y))\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \tag{4-102}$$

式(4-101)也包含了式(4-99)的情况，因此适用于任何一对一映射。最后，由于 $y=g(x)$ 的微分更容易计算，通常使用下式：

$$f_Y(y)=\frac{f_X(x)}{|g'(x)|}\bigg|_{x=g^{-1}(y)} \tag{4-103}$$

其中 $g'(x)$ 是函数关于 x 的微分。

上面的间接方法也可以用于任意函数，但使用下面的直接方法会更加简单。

- 任意 $g(X)$ 的直接方法。假设任何区域中函数的导数都不为 0。得到类似式(4-103)的变化形式，当 $g'(x)=0$ 时不能直接计算。稍后讨论函数为定值的情况。

图 4-25 给出了非单调函数的例子，从图中可以看出，每个 y 的值对应两个不同的 x 值(除了 $x=0$ 处)。尽管我们分析的是简单的均匀分布随机变量，但是可以很容易将结果推广到从 x 到 y 的任意数目(包括可数无限)映射区间。从小区间 $[y,y+\mathrm{d}y]$ 上 y 的概率开始：

$$\begin{aligned}P(y\in[y,y+\mathrm{d}y])&=P(x\in[x_1,x_1+\mathrm{d}x_1]\cup x\in[x_2,x_2+\mathrm{d}x_2])\\&=P(x\in[x_1,x_1+\mathrm{d}x_1])+P(x\in[x_2,x_2+\mathrm{d}x_2])\end{aligned} \tag{4-104}$$

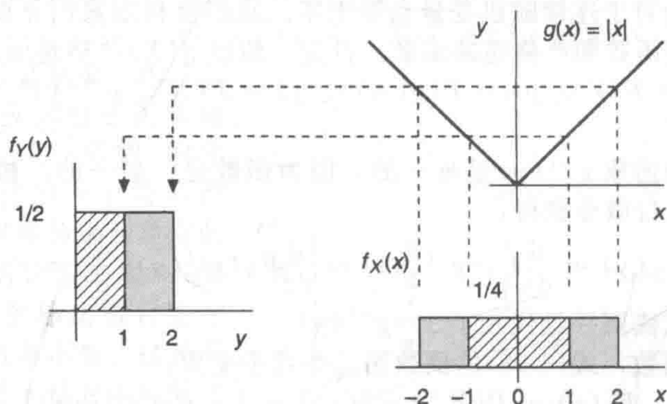


图 4-25 连续随机变量的非单调函数映射。两个阴影区域标出了映射到同一 Y 值的 X 区间(多对一函数)

由于区间选择时保证了不相交，所以第二个等式成立。这些概率可以通过下列矩形面积近似：

$$f_Y(y)|\mathrm{d}y|\approx f_X(x_1)|\mathrm{d}x_1|+f_X(x_2)|\mathrm{d}x_2| \tag{4-105}$$

其中，区间要足够小以保证 $|\mathrm{d}x_1|\approx|\mathrm{d}x_2|\triangleq|\mathrm{d}x|$。在微分学里，$\{x_1,x_2,y\}$ 可能是不同区间的中心点，但这并不重要，因为每个区间的大小最终将趋于零。绝对值 $|\mathrm{d}x_n|$ 确保式(4-104)中的 $\mathrm{d}x_1$、$\mathrm{d}x_2$ 为负值情况下(当然，概率密度函数总是非负的)，面积为正。最后，将这个结果除以 $\mathrm{d}y$，代入逆象 $x=g^{-1}(y)$，将此方法扩展到函数的 N 个不同区间，得到：

$$f_Y(y)=\sum_{n=1}^{N}\frac{f_X(x_n)}{|g'(x_n)|}\bigg|_{x_n=g_n^{-1}(y)} \tag{4-106}$$

其中，$g_n(x)$ 的下标表示对应 x_n 的函数的第 n 个区域。

从图 4-26 所示的例子可以看出，函数 $g(x)$ 的区域数目很容易确定。绘制通过函数的水平线可以确定在多个 y 值中，

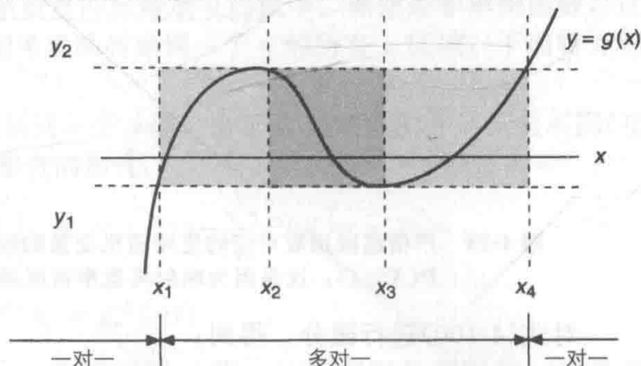


图 4-26 X 到 Y 映射的非单调函数的区间由与 $g(x)$ 相交的水平线确定

究竟在什么地方它们与$y=g(X)$相交。在这个例子中，有三个不同的区域：(ⅰ) (x_4,∞) 一对一映射到 (y_2,∞)，(ⅱ) $(-\infty,x_1]$ 一对一映射到 $(-\infty,y_1]$，(ⅲ) $(x_1,x_4]$ 的三个子区域映射到相同的 $(y_1,y_2]$ 区间。图中的这些子区域分别为 $\{(x_1,x_2],(x_2,x_3],(x_3,x_4]\}$（注意，因为假定 $g(X)$连续，所以哪个区间的终点闭合并不重要）。由于每个子区域的函数形状不同(这不是必要条件)，所以式(4-106)的 $g_n(\cdot)$有个下标。通过水平线可以容易地看出函数从递增变为递减的位置，反之亦然，并且可以看出哪里有多对一映射。

绘制通过函数的水平线相当于找到不同 y 值下 $g(x)-y=0$ 的根。例如，$y=g(x)=x^2-1$，相当于求 $x^2-y-1=0$ 的根，结果为 $x=\pm\sqrt{y+1}$。两个 x 值映射到相同的 y 值($y=0$ 除外)，这种情况下式(4-106)的 $N=2$。因此一般情况下，将式(4-106)变成下面的简化形式：

$$f_Y(y)=\sum_{n=1}^{N}\frac{f_X(x_n)}{|g'(x_n)|} \tag{4-107}$$

$\{x_n\}$为 $g(x)-y=0$ 的根，$n=1,\cdots,N$。

例 4-10 设 X 是参数为$\{\mu=0,\ \alpha\}$的拉普拉斯随机变量，考虑映射 $Y=X^2$，也就是 $N=2$。由于 $g'(x)=2x$，所以

$$f_Y(y)=\frac{1}{2\alpha}\sum_{n=1}^{2}\frac{1}{|2x_n|}\exp(-|x_n|/\alpha) \tag{4-108}$$

值域为$\mathcal{R}^+$。$g(x)-y=0$ 的根为 $x_n=\pm\sqrt{y}$，因此

$$\begin{aligned} f_Y(y) &= \frac{1}{4\alpha\sqrt{y}}[\exp(-|\sqrt{y}|/\alpha)+\exp(-|-\sqrt{y}|/\alpha)] \\ &= \frac{1}{2\alpha\sqrt{y}}\exp(-\sqrt{y}/\alpha)I_{[0,\infty)}(y) \end{aligned} \tag{4-109}$$

其中，最终结果里忽略了绝对值。图 4-27 给出了 $\alpha=1$ 的概率密度函数，因为$\sqrt{y}$在分母中，所以概率密度函数(虚线)在原点处增大到无穷大。其根本原因在于 $g'(x)$在 $x=0$ 处趋近于零。(通过高斯随机变量的平方获得的自由度为 1 的卡方随机变量，在原点有相似的渐近行为。)

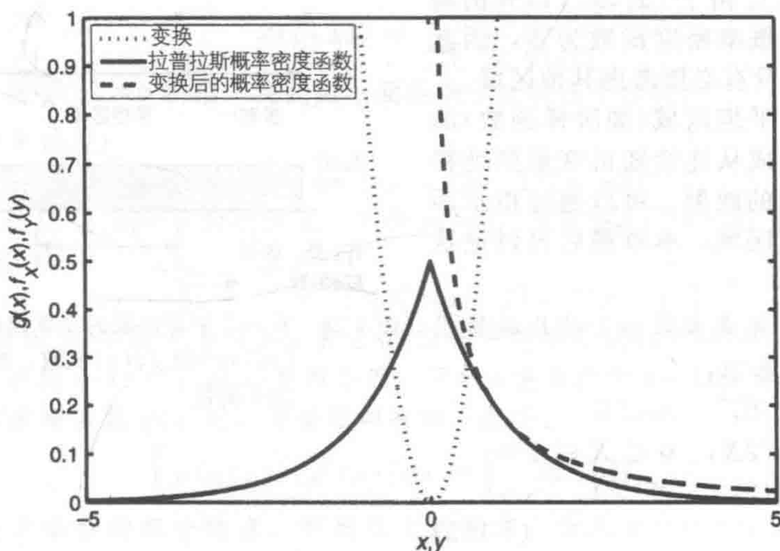


图 4-27　例 4-10 的概率密度函线。对于拉普拉斯概率密度函数，$\alpha=1$，变换 $Y=X^2$

对于任意的函数，很难利用间接(累积分布函数)的方法，因为跟踪映射到相同的 y 值的所有不同 x 轴区域是比较繁琐的。然而，对于相对简单的函数它却很有用，比如例 4-11 中的函数。

例 4-11 对于变换 $Y=X^2$，我们考察下面的概率：

$$P(Y \leqslant y)=P(X^2 \leqslant y)=P(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) \tag{4-110}$$

写成 X 的累积分布函数形式：

$$F_Y(y)=F_X(\sqrt{y})-F_X(-\sqrt{y}) \tag{4-111}$$

将上式对 y 微分，得到通用结果：

$$f_Y(y)=\frac{1}{2\sqrt{y}}[f_X(\sqrt{y})+f_X(-\sqrt{y})] \tag{4-112}$$

代入拉普拉斯随机变量的概率密度函数得到例 4-10中的式(4-109)。◀

- **有平坦区间的函数**。接下来，我们介绍当函数有一个或多个区间的导数为 0 时，如何处理变换。图 4-28 给出了函数有一个上述情况的例子。观察发现该函数可以被划分成多个区间：（ⅰ）严格递增或严格递减的区域，（ⅱ）$g'(x)=0$ 的平坦区域。第一种情况对应的区域由式(4-106)进行处理。接下来考虑图 4-28 中 $[x_2, x_3]$ 对应的平坦区域(浅色阴影表示)，观察发现：

$$P(Y=y)=P(X \in [x_2, x_3])=\int_{x_2}^{x_3} f_X(x)\mathrm{d}x \triangleq P_Y(y) \tag{4-113}$$

由于 $x_3>x_2$，所以上式不为 0。$f_X(x)$下的面积为 $P(Y=y)$的概率，我们用速记符号 $P_Y(y)$(它不是 $p_Y[y]$的概率质量函数)表示。将这个结果与式(4-106)结合，得到 Y 的混合概率密度函数，假设有 M 个平坦区域：

$$f_Y(y)=\sum_{n=1}^{N} \frac{f_X(x_n)}{|g'(x_n)|}\bigg|_{x_n=g_n^{-1}(y)}+\sum_{m=1}^{M} P_Y(y_m)\delta(y-y_m) \tag{4-114}$$

其中 $\delta(y)$是狄拉克 δ 函数。注意，图 4-28 中略小于 y_2 的一条横线，在区间 $x \in [x_1, x_2] \cup [x_3, x_4]$ 与 $g(x)$相交三次，这些区间映射到 $[y_1, y_2]$。区间 $[x_4, x_5]$ 一对一映射到 $[y_2, y_3]$。由于 $[x_1, x_5]$ 以外的均匀分布的概率密度函数为零，因此对此变换没有必要考虑其他区域。

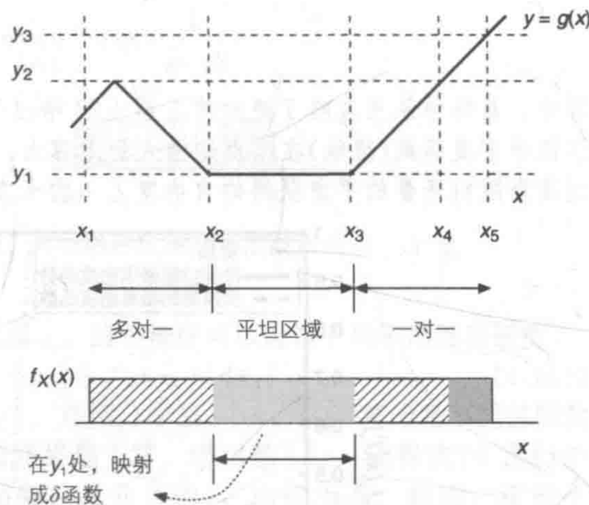


图 4-28 有一个平坦区域的非单调函数的例子 $x \in [x_2, x_3]$ 的 $f_X(x)$区域，映射到 $y=y_1$ 处为 δ 函数

函数在只有平坦区域(如阶梯函数)的极端情况下，变成从连续随机变量到纯粹的离散随机变量的映射。可以通过指示函数表示多个平坦区域，本章稍后再讨论这种映射。

例 4-12 设 X 是参数为 λ 的指数随机变量，它的映射如下：

$$g(X)=\begin{cases}0, & x<0 \\ 2X, & 0 \leqslant X \leqslant 1 \\ 2, & X>1\end{cases} \tag{4-115}$$

产生了混合随机变量。对线性区域采用间接法，当 $X \in [0,1]$ 时，$Y=g(X)$的累积分布函数为：

$$F_Y(y)=P(2X \leqslant y)=P(X \leqslant y/2)=F_X(y/2) \tag{4-116}$$

对上式关于 y 进行微分，得到：

$$\frac{\mathrm{d}}{\mathrm{d}y}F_X(y/2)=(1/2)f_X(y/2)=(\lambda/2)\exp(-\lambda y/2)I_{[0,2]}(y) \tag{4-117}$$

这里采用指示函数说明 $f_Y(y)$ 的连续部分。离散部分通过下式获得：

$$P(Y=2)=P(X>1)=\lambda\int_1^{\infty}\exp(-\lambda x)\mathrm{d}x=\exp(-\lambda) \tag{4-118}$$

因此，Y 的整体概率密度函数为：

$$f_Y(y)=(\lambda/2)\exp(-\lambda y/2)I_{[0,2]}(y)+\exp(-\lambda)\delta(y-2) \tag{4-119}$$

它直接表明这个概率密度函数具有单位面积。图 4-29给出了变换和两个随机变量的概率密度函数。

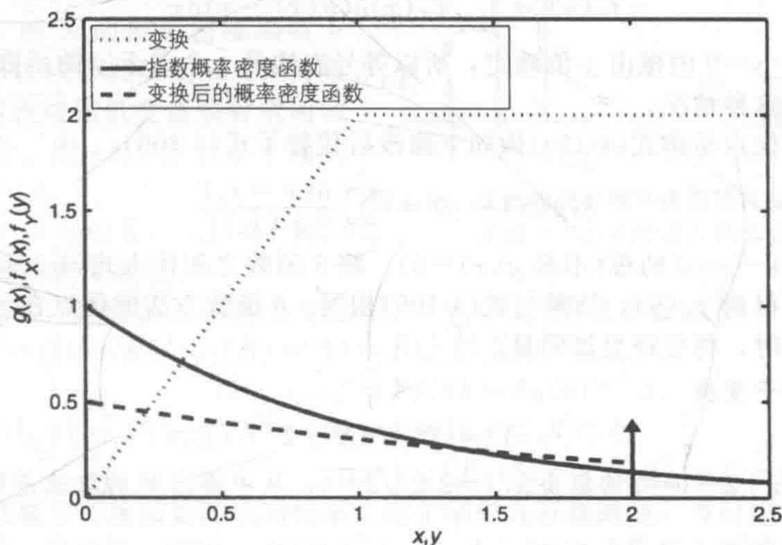


图 4-29　例 4-12 中随机变量变换的概率密度函数，其中 X 服从 $\lambda=1$ 的指数分布 ◀

- 任意 $g(X)$ 的 δ 函数法。当连续函数是平滑函数 $g(x)$（见附录 B）的组合时，有可能利用下面的狄拉克 δ 函数的性质改善对连续函数的变换过程：

$$\delta(g(x))=\sum_{n=1}^{N}\frac{\delta(x-x_n)}{|g'(x_n)|} \tag{4-120}$$

其中 $\{x_n\}$ 是 $g(x)=0$ 的根。我们用下面的例子证明这个性质。

例 4-13 令变换为：

$$g(x)=x^2-4x+3=(x-1)(x-3) \tag{4-121}$$

其中 $g'(x)=2x-4$。$g(x)=0$ 的根为 $x_1=1$ 和 $x_2=3$，因此 $g'(1)=-2$，$g'(3)=2$。利用式(4-120)，得出：

$$\delta(x^2-3x+2)=(1/2)[\delta(x-3)-\delta(x-1)] \tag{4-122}$$

因为 $\delta(x)$ 仅当 $x=0$ 时是非零的，所以我们感兴趣的是式(4-120)中 δ 函数的自变量 $g(x)=0$ 的根。在例 4-13 中，$g(x)$ 有两个根，因此 δ 函数位于 $x=1$ 和 3 处。利用微积分学，当变量变成单调函数 $g(x)$ 时，积分区间有如下改变：

$$\int_a^b h(g(x))|g'(x)|\mathrm{d}x=\int_{g(a)}^{g(b)}h(x)\mathrm{d}x \tag{4-123}$$

回顾狄拉克 δ 函数的积分特点，利用以上的结果，代入 $a=-\infty$，$b=\infty$，$h(x)=\delta(x)$，得到：

$$\int_{-\infty}^{\infty}\delta(g(x))|g'(x)|\mathrm{d}x=\int_{-\infty}^{\infty}\delta(x)\mathrm{d}x \tag{4-124}$$

重新整理被积函数，得到：

$$\delta(g(x))=\frac{1}{|g'(x)|}\delta(x) \tag{4-125}$$

式(4-120)是非单调函数 $g(x)=0$ 有 N 个根时，式(4-125)的一个扩展。◀

比较式(4-120)和式(4-106)，式(4-106)用于确定 $Y=g(X)$ 的概率密度函数，我们认识到了它们的相似性。事实上，利用狄拉克 δ 函数筛选特性：

$$\int_{-\infty}^{\infty} f_X(x)\delta(g(x))\mathrm{d}x=\int_{-\infty}^{\infty} f_X(x)\sum_{n=1}^{N}\frac{\delta(x-x_n)}{|g'(x_n)|}\mathrm{d}x=\sum_{n=1}^{N}\frac{f_X(x_n)}{|g'(x_n)|} \tag{4-126}$$

其中$\{x_n\}$是 $g(x)=0$ 的根。为了获得变换后的 $f_Y(y)$，像式(4-107)后面讨论的那样，$\{x_n\}$必须是 $g(x)-y=0$ 的根。因此，我们可以将式(4-126)中的 $\delta(g(x))$ 替换为 $\delta(g(x)-y)$，得到最终的结果：

$$f_Y(y)=\int_{-\infty}^{\infty} f_X(x)\delta(g(x)-y)\mathrm{d}x \tag{4-127}$$

由于 $g(x)-y=0$ 的根由 y 值确定，所以等号右边是一个关于 y 的函数，式(4-107)中初始表达式就是这种情况。

这种方法的优点是将式(4-120)做如下修改后代替了式(4-106)：

$$\delta(g(x)-y)=\sum_{n=1}^{N}\frac{\delta(x-x_n)}{|g'(x_n)|} \tag{4-128}$$

其中$\{x_n\}$是 $g(x)-y=0$ 的根(不是 $g(x)=0$)。将 δ 函数之和代入式(4-127)，再利用 δ 函数的筛选特性，得到 $f_Y(y)$，结果与式(4-106)相同。δ 函数方法的优点在分析两个或多个随机变量的变换时，将变得更加明显。

例 4-14 对于变换

$$Y=X^2-4X+2 \tag{4-129}$$

$g(x)-y=x^2-4x+2-y=0$ 的根为 $x_{1,2}=2\pm\sqrt{2+y}$，从中得出 Y 的取值范围是$[-2,\infty)$。由于

$$g'(x_1)=2\sqrt{2+y},\quad g'(x_2)=-2\sqrt{2+y} \tag{4-130}$$

Y 的概率密度函数为：

$$f_Y(y)=\frac{1}{2\sqrt{2+y}}[f_X(2+\sqrt{2+y})+f_X(2-\sqrt{2+y})] \tag{4-131}$$

图 4-30 给出了 X 为标准高斯随机变量时的 $f_Y(y)$。

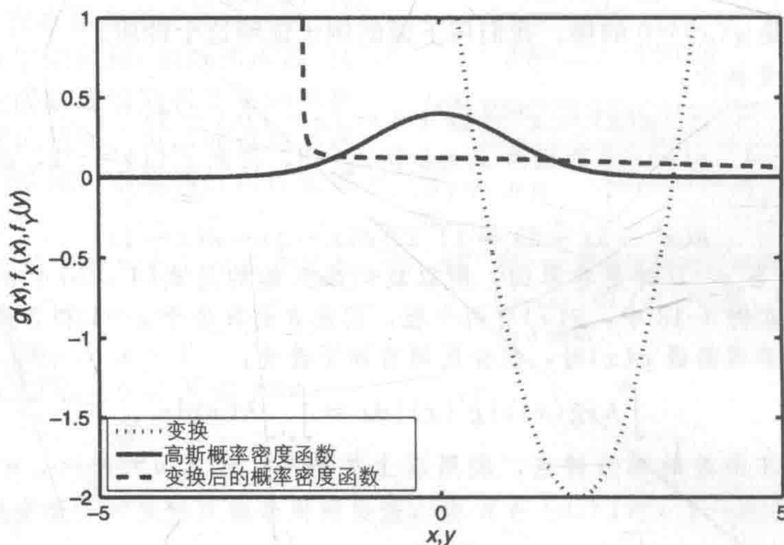


图 4-30 例 4-14 中随机变量变换的概率密度函数，其中 X 服从标准高斯分布 ◀

- **指示函数法**。许多应用会将连续信号变换为严格离散的幅度信号。例如，在数字通信中，用符号函数做决策器将一个连续的波形变为两个值：±1。同样，单位阶跃

函数可将一个连续的波形映射为 $\{0,1\}$。第 5 章将介绍如何利用阶梯函数集生成的离散随机变量序列来近似连续随机变量。

将 $a<b$ 条件下的指示函数 $I_{[a,b]}(x)$ 用于变换是比较方便的，将它的自变量用随机变量 X 取代：

$$Y = I_{[a,b]}(X) \tag{4-132}$$

连续随机变量 X(具有任意概率密度函数)的映射产生了输出为 $\{0,1\}$ 的伯努利随机变量 Y。为了得到 $f_Y(y)$，我们只需确定映射到 0 的 X 的概率密度函数下的面积，剩下的区域必然映射到 1。符号函数是另一种将连续随机变量映射成离散随机变量的方法。图 4-31 显示了下面的映射：

$$Y = \text{sgn}(X) \tag{4-133}$$

其中 X 有均匀概率密度函数。

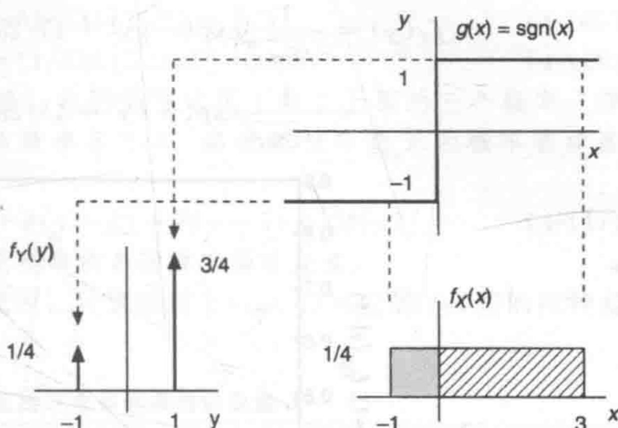


图 4-31　连续随机变量到离散随机变量的映射。正负区间 $f_X(x)$ 下的面积映射成 $f_Y(y)$ 的 δ 函数

回到指示函数，并假设 Y 的输出为 $\{0,1\}$，$f_Y(y)$ 的推导过程如下：

$$\begin{aligned} P(Y=1) = P(I_{[a,b]}(X)=1) &= P(a \leqslant X \leqslant b) \\ &= F_X(b) - F_X(\text{a}) = p \end{aligned} \tag{4-134}$$

$$\begin{aligned} P(Y=0) = P(I_{[a,b]}(X)=0) &= 1 - P(a \leqslant X \leqslant b) \\ &= 1 - [F_X(b) - F_X(a)] = 1 - p \triangleq q \end{aligned} \tag{4-135}$$

其中 p 是伯努利概率密度函数的成功概率。这个结果具有普遍性，可以用来表示从连续到离散随机变量的一类变换。例如，令 $[a,b]=[0,\infty)$，指示函数和上面的表达式可以用来实现单位阶跃函数。对于符号函数，区间也是 $[a,b]=[0,\infty)$，但式(4-135)中的结果由 0 变为 -1。对于所有概率密度函数关于原点对称的随机变量 X，由此产生的对称伯努利随机变量 Y 必然具有等概率的输出：$p=q=1/2$。

例 4-15　如前所述，数字通信中常使用符号函数来检测离散时刻传输的是 1 还是 -1。具有加性噪声的系统的简化框图如图 4-32 所示。假设符号为 ± 1 的信号每 T 秒随机传送一次，各离散时刻对应一个对称的伯努利随机变量。假定加性噪声 W 是参数为 $\{\mu=0,\sigma\}$ 的高斯随机变量。虽然 X 是一个离散随机变量，但接收信号 Y 是连续的，并且由于加性噪声的存在，它可以是 $\mathcal{R}$ 中的任意值。在随后的一节中可以得出，当 X 和 W 相互独立时，$Y=X+W$ 的概率密度函数由以下卷积得到：



图 4-32　二值通信系统的简化框图

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(v) f_W(y-v)\,\mathrm{d}v \tag{4-136}$$

由于 $f_X(x)$ 是一个离散随机变量，它的概率密度函数是狄拉克 δ 函数，因此卷积很容易计算：

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} (1/2)[\delta(v-1)+\delta(v+1)] \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(y-v)^2/2\sigma^2)\,\mathrm{d}v \\ &= (1/2)\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(y-1)^2/2\sigma^2) + \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(y+1)^2/2\sigma^2)\right] \\ &= (1/2)[f_{Y|X}(y|-1) + f_{Y|X}(y|+1)] \end{aligned} \tag{4-137}$$

这里利用了狄拉克 δ 函数的筛选特性。这个结果是两个条件高斯概率密度函数的和，一个

中点为 1，另一个中点为 −1。Y 的概率密度函数和两个条件概率密度函数如图 4-33 所示。通过提取式(4-137)的公共指数项，结果可写为包含双曲余弦函数的乘积形式：

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{8\pi\sigma^2}}\exp(-(y^2+1)/2\sigma^2)[\exp(-y/\sigma^2)+\exp(y/\sigma^2)] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(y^2+1)/2\sigma^2)\cosh(y/\sigma^2) \end{aligned} \tag{4-138}$$

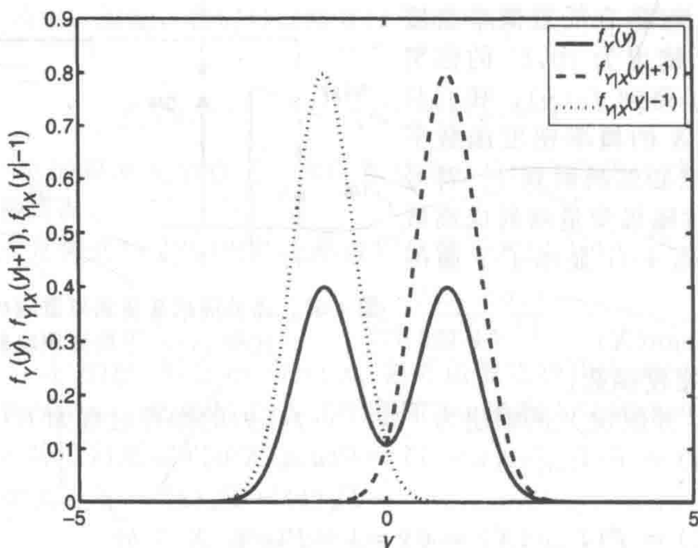


图 4-33 例 4-15 中二值通信系统的接收信号随机变量 Y 的双峰概率密度函数以及条件概率密度函数 $f_{Y|X}(y|-1)$、$f_{Y|X}(y|+1)$

在 $x=\pm1$ 处有两个峰。符号函数检测器通过以下方法"决定"传输哪个符号

$$\hat{X}=\operatorname{sgn}(Y) \tag{4-139}$$

其中$\hat{X}$表示 X 的估计(我们任意假定 $Y=0$ 时$\hat{X}=1$)。从对称性可知，当传输信号为等概率时，$P(\hat{X}=1)=P(\hat{X}=-1)=1/2$。这并不意味着将选择出正确的 X 值。如果 $X=-1$ 发送时，噪声值超过 1，符号函数检测器会错误地认为 1 在传输。在给定条件概率 $P(\hat{X}=1|X=-1)=P(\hat{X}=-1|X=1)$后，错误概率是衡量接收器性能的指标。第 10 章将介绍这些概率取决于 W 的方差：较大的噪声方差将产生更大的错误概率。◀

使用指示函数的变换方法可以概括如下：

$$Y=\sum_{m=1}^{M} y_m I_{[a_m,b_m]}(X) \tag{4-140}$$

其中区间 $[a_m,b_m]$ 表示 X 的取值范围，则 $g(X)$映射到 y_m。相应的概率为：

$$P(Y=y_m)=P(a_m\leqslant X\leqslant b_m)=\int_{-\infty}^{\infty} f_X(x)I_{[a_m,b_m]}(x)\mathrm{d}x \triangleq P_Y(y_m) \tag{4-141}$$

书写方式与前面混合随机变量变换的相同。因此，Y 的概率密度函数由式(4-114)的离散部分获得：

$$f_Y(y)=\sum_{m=1}^{M} P_Y(y_m)\delta(y-y_m) \tag{4-142}$$

这里认为对应$\{P_Y(y_m)\}$的区间 $\{[a_m,b_m]\}$ 互不重叠，并且它们有可能形成 x 轴的一个分割(partition)。

例 4-16 设 X 服从$[0,3]$上的均匀分布，考虑映射：

$$Y=g(X)=2I_{[0,1/2]}(X)+I_{[1,2]}(X)-I_{[5/2,3]}(X) \tag{4-143}$$

它不是$[0,3]$的一个分割。这意味着 X 的一些区间映射到 0，从而 Y 的输出为 $\{-1,0,1,$

2}。由于 X 是均匀分布的，因此勒贝格测度可以用来计算概率：

$$P(0 \leqslant X \leqslant 1/2) = (1/3)L[0,1/2] = 1/6 \tag{4-144}$$

$$P(1 \leqslant X \leqslant 2) = (1/3)L[1,2] = 1/3 \tag{4-145}$$

$$P(5/2 \leqslant X \leqslant 3) = (1/3)L[5/2,3] = 1/6 \tag{4-146}$$

其中 1/3 为 $f_X(x)$ 的高度。使用勒贝格测度，或更简单地用 1 减去上面的三个概率，得到 X 在两个区间 $\{[1/2,1],[2,5/2]\}$ 上的概率是 1/3。离散随机变量 Y 的概率密度函数为：

$$f_Y(y) = (1/6)[2\delta(y-1) + \delta(y-2) - \delta(y+1) + 2\delta(y)] \tag{4-147}$$

其中确定了 Y 的输出不能与给定了离散输出概率的 δ 函数的面积混淆。◀

表 4-2 总结了当 X 是一个连续随机变量时，计算映射 $Y=g(X)$ 对应的 $f_Y(y)$ 的四种基本变换方法。

表 4-2　$Y=g(X)$ 的变换方法和连续随机变量 X

累积分布函数(间接)方法
首先计算 Y 的累积分布函数：$F_Y(y)$(假设 $g(X)$ 单调)
将 $F_Y(y)$ 写成 $F_X(\cdot)$ 的形式：$P(Y\leqslant y)=P(X\leqslant g^{-1}(y))=F_X(g^{-1}(y))$
对 y 微分 $f_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}F_X(g^{-1}(y))=f_X(g^{-1}(y))\frac{\mathrm{d}x}{\mathrm{d}y}$
概率密度函数(直接)方法
计算 $g(x)=0$ 的所有根，并且将它们写成 y 的形式：$x_n=g^{-1}(y)$
计算 $g'(x)$ 并且对每个 x_n 计算 $f_X(x)/g'(x)$ 的比值
将所有根相加：$f_Y(y)=\sum_{n=1}^{N} f_X(x_n)/\lvert g'(x_n)\rvert$
代入 $x_n=g^{-1}(y)$
δ 函数方法
计算 $g(x)-y=0$ 的所有根
利用 $\delta(g(x)-y)=\sum_{n=1}^{N}\delta(x-x_n)/\lvert g'(x_n)\rvert$，重写 $\delta(g(x)-y)$
积分并且利用筛选特性：$f_Y(y)=\int_{-\infty}^{\infty} f_X(x)\delta(g(x)-y)\mathrm{d}x$
指示函数方法
将 $g(X)$ 表示成加权指示函数之和的形式
确定 $P_Y(y_m)=P(a_m\leqslant X\leqslant b_m)$ 的概率
对 δ 函数加权求和：$f_Y(y)=\sum_{m=1}^{M} P_Y(y_m)\delta(y-y_m)$

4.9　两个随机变量的重要函数

本节将考虑两个独立随机变量到一个随机变量的变换。除了首先讨论的随机变量之和外，我们仅考虑两个随机变量都连续的情况。对于离散随机变量的情况，或者一个连续另一个离散(如例 4-15)的情况，都可参照连续方法。

4.9.1　求和：$Z=X+Y$

累积分布函数(间接)方法可用于导出当 X 和 Y 是独立的连续随机变量时，Z 的概率密度函数。首先计算联合累积分布函数 $F_{X,Y}(x,y)$，Z 的累积分布函数可以写成以下二重积分的形式：

$$F_Z(z) = P(X+Y \leqslant z) = \iint_{x+y\leqslant z} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{4-148}$$

其中在最后一个表达式里，将 $x+y\leqslant z$ 替换为 $x\leqslant z-y$，并用它来限制内积分的积分范围。关于 z 进行微分得到：

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z-y,y)\mathrm{d}y \tag{4-149}$$

对相关的 X 和 Y 上式通用，当它们相互独立时，联合概率密度函数可分开计算，得到最终结果：

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\mathrm{d}y \tag{4-150}$$

我们看到它是两个边缘概率密度函数的卷积：$f_Z(z)=f_X(x)*f_Y(y)$。

现在假设 X 连续 Y 离散，式(4-150)变为：

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(z-y)\sum_{n=-\infty}^{\infty} p_Y[n]\delta(y-n)\mathrm{d}y \tag{4-151}$$

这里代入了基于狄拉克 δ 函数的 Y 的概率密度函数。利用筛选特性，卷积变为：

$$f_Z(z) = \sum_{n=-\infty}^{\infty} f_X(z-n)p_Y[n] \tag{4-152}$$

这个表达式是 X 的连续概率密度函数变换后的加权和。加权值由 Y 的概率质量函数给定，以确保 $f_Z(z)$具有单位面积。它是本章稍后将要学习的“混合”的一个例子。

最后，当 X 和 Y 都离散时，先采用类似式(4-148)的方法，但等式为：

$$P(Z=z) = P(X+Y=z) = \sum_{x+y=z} p_{X,Y}[x,y] \tag{4-153}$$

这是一个一对一的求和，当 x 变化时，等式 $x+y=z$ 使得 $y=z-x$。假设独立并将概率质量函数里的变量 y 替换为 $z-x$，得到最终结果，这是一个离散卷积：

$$p_Z[z] = \sum_{x=-\infty}^{\infty} p_X[x]p_Y[z-x] \tag{4-154}$$

例 4-17 设 X 和 Y 是独立同分布的参数为 λ 的指数分布随机变量。$Z=X+Y$ 的概率密度函数是：

$$f_Z(z) = \lambda^2\int_{-\infty}^{\infty} \exp(-\lambda(z-y))\exp(-\lambda y)I_{[0,\infty)}(z-y)I_{[0,\infty)}(y)\mathrm{d}y \tag{4-155}$$

指示函数用来跟踪每个函数的区间，当被积函数里的参数变化时，它是非常重要的。积分区间为 $[0,\infty)\cap(-\infty,z]=[0,z]$，得到：

$$f_Z(z) = \lambda^2\exp(-\lambda z)\int_0^z \mathrm{d}y = \lambda^2 z\exp(-\lambda z)I_{[0,\infty)}(z) \tag{4-156}$$

这是参数为$\{\lambda, r=2\}$的埃尔朗(伽马)分布。◀

例 4-18 设 X_1 和 X_2 是参数为$\{N, p\}$的独立同分布二项分布随机变量。利用式(4-154)，得出 $Y=X_1+X_2$ 的概率质量函数：

$$\begin{aligned} p_Y[y] &= \sum_{x_1=-\infty}^{\infty}\binom{N}{x_1}p^{x_1}q^{N-x_1}\binom{N}{y-x_1}p^{y-x_1}q^{N-y+x_1}I_{\{0,\cdots,N\}}[x_1]I_{\{0,\cdots,N\}}[y-x_1] \\ &= p^y q^{2N-y}\sum_{x_1=0}^{y}\binom{N}{x_1}\binom{N}{y-x_1}I_{\{0,\cdots,2N\}}[y] \end{aligned} \tag{4-157}$$

离散指示函数用于指定正确的求和限。可以证明(见习题 4-30)，最后一个表达式的求和可以简化为$\binom{2N}{y}$。因此结果是参数为$\{2N,p\}$的二项分布的概率质量函数：

$$p_Y[y] = \binom{2N}{y}p^y q^{2N-y}I_{\{0,\cdots,2N\}}[y] \tag{4-158}$$ ◀

4.9.2 差：$Z=X-Y$

使用类似 4.9.1 节的方法，差的累积分布函数为：

$$F_Z(z)=P(X-Y\leqslant z)=\iint_{x-y\leqslant z} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{z+y} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{4-159}$$

关于 z 进行微分，得到

$$f_Z(z)=\int_{-\infty}^{\infty} f_{X,Y}(z+y,y)\mathrm{d}y=\int_{-\infty}^{\infty} f_X(z+y)f_Y(y)\mathrm{d}y \tag{4-160}$$

最后一个表达式是利用随机变量独立得出的。这个积分是两个概率密度函数的确定的互相关函数：$f_Z(z)=f_Y(y)*f_X(x)$（见附录 C 中的定义）。不像卷积，如我们预计的那样，它不是对称的，因为 $Z=X-Y$ 关于 X 和 Y 不对称：

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(x-z)\mathrm{d}x \tag{4-161}$$

离散随机变量的相应结果是：

$$p_Z[z]=\sum_{y=-\infty}^{\infty} p_X[z+y]p_Y[y]=\sum_{x=-\infty}^{\infty} p_X[x]p_Y[x-z] \tag{4-162}$$

例 4-19 设 X 和 Y 是参数为 p 的独立同分布伯努利随机变量，因此 $Z=X-Y$ 的输出为 $\{-1,0,1\}$。其概率质量函数为：

$$p_Z[z]=\sum_{y=-\infty}^{\infty} p^{z+y}q^{1-y-z}p^yq^{1-y}I_{\{0,1\}}[z+y]I_{\{0,1\}}[y]$$
$$=q^2(p/q)^z\sum_{y=0}^{1}(p/q)^{2y}I_{\{0,1\}}[z+y] \tag{4-163}$$

其中第二个指示函数给出了求和范围是 $\{0,1\}$。给定 z 时，第一个指示函数按照如下方式限制 x 的值：$z=-1\Rightarrow y=1$，$z=0\Rightarrow y=\{0,1)$ 以及 $z=1\Rightarrow y=0$。从而

$$p_Z[z]=q^2[(p/q)^{-1}(p/q)^2\delta[z+1]+(1+(p/q)^2)\delta[z]+(p/q)\delta[z-1]]$$
$$=pq\delta[z+1]+(q^2+p^2)\delta[z]+pq\delta[z-1] \tag{4-164}$$

这是一个对称的（偶）概率质量函数，不与任何标准参数分布对应。◀

4.9.3 积：$Z=XY$

考虑独立随机变量 X 和 Y 相乘的情况。再次使用累积分布函数方法：

$$F_Z(z)=P(XY\leqslant z)=\iint_{xy\leqslant z} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$
$$=\int_{-\infty}^{0}\int_{z/y}^{\infty} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y+\int_{0}^{\infty}\int_{-\infty}^{z/y} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{4-165}$$

因为 y 可能为负，这种情况将改变相除时的不等号方向，所以外积分必须分成两个区域：$y\in(-\infty,0]$ 和 $y\in[0,\infty)$。确定对应内积分区间的不等式分别为 $x\geqslant z/y$ 和 $x\leqslant z/y$。对两部分关于 z 进行微分，利用链式法则得到：

$$f_Z(z)=-\int_{-\infty}^{0}(1/y)f_{X,Y}(z/y,y)\mathrm{d}y+\int_{0}^{\infty}(1/y)f_{X,Y}(z/y,y)\mathrm{d}y$$
$$=\int_{-\infty}^{\infty}\frac{1}{|y|}f_{X,Y}(z/y,y)\mathrm{d}y \tag{4-166}$$

当 X 和 Y 独立时，最终的结果是：

$$f_Z(z)=\int_{-\infty}^{\infty}\frac{1}{|y|}f_X(z/y)f_Y(y)\mathrm{d}y \tag{4-167}$$

对于离散随机变量，表达式是：

$$p_Z[z] = p(XY = z) = \sum_x \sum_{y=z/x} p_X[x] p_Y[y] \tag{4-168}$$

这里认为 z/x 是整数，一定要仔细考虑离散情况下乘积中的所有项。

例 4-20 设 X 是一个参数为 p 的几何分布随机变量，Y 服从伯努利分布，参数同样为 p。乘积 $Z=XY$ 的值域是$\mathcal{Z}^+$，与 X 相同。因为 Y 服从伯努利分布且 $Y=0$ 出现的概率为 $q \triangleq 1-p$，我们预计 Z 将在 $z=0$ 处有更大的概率质量(与几何随机变量相比)。计算概率密度函数时需要分别考虑 $Z=0$ 和 $Z\neq 0$ 的情况：

$$\begin{aligned} P(Z=0) &= P(X=0 \text{ 或 } Y=0) \\ &= P(X=0)+P(Y=0)-P(X=0)P(Y=0) \\ &= p+q-pq = 1-pq \end{aligned} \tag{4-169}$$

其中第二行利用了独立性。对于 $Z>0$：

$$\begin{aligned} P(Z=z) &= P(X=z \text{ 且 } Y=1) \\ &= p(X=z)P(Y=1) = p^2 q^z \end{aligned} \tag{4-170}$$

结合这两种结果，得到以下概率质量函数：

$$p_Z[z] = (1-pq)\delta[z] + p^2 \sum_{n=1}^{\infty} q^n \delta[z-n] \tag{4-171}$$

需要注意的是，可以加上或减去 $p^2\delta[z]$，因此求和中包括 $n=0$：

$$p_Z[z] = q\delta[z] + p \sum_{n=0}^{\infty} pq^n \delta[z-n] \tag{4-172}$$

这是一个由 p 加权的几何分布概率密度函数以及只有一个输出 0 的“随机”变量的概率质量函数 $\delta[z]$的混合。

接下来，利用式(4-168)，通过对 y 求和，并且在第二个求和中代入 y 的值，推导得到：

$$p_Z[z] = \sum_y \sum_{x=z/y} pq^x p^y q^{1-y} = \sum_{x:x\cdot 0=z} pq^x q + \sum_{x=z} pq^x p \tag{4-173}$$

第一个求和只用于 $z=0$，这是因为乘积项中 $y=0$，所以对所有 x 值求和：

$$\sum_{x:x\cdot 0=0} pq^x q = pq \sum_{x=0}^{\infty} q^x = pq/(1-q) = q \tag{4-174}$$

第二个求和仅关于每个 z 值进行：

$$\sum_{x:x\cdot 1=z} pq^x p = p^2 \sum_{x=z} q^x = p^2 q^z I_{\mathcal{Z}^+}[z] \tag{4-175}$$

结合这两个结果得出了式(4-172)中的概率质量函数。◀

例 4-21 设 X 是参数为λ 的指数分布随机变量，Y 服从参数为 p 的伯努利分布。直观地说，我们预计乘积 $Z=XY$ 是一个在 $Z=0$ 处概率质量非 0 的混合随机变量

$$P(Z=0) = P(XY=0) = P(Y=0) = q \tag{4-176}$$

首先计算累积分布函数：

$$\begin{aligned} P(Z\leqslant z) &= P(Z\leqslant z|Y=0)P(Y=0)+P(Z\leqslant z|Y=1)P(Y=1) \\ &= qu(z)+pP(Z\leqslant z|Y=1) \end{aligned} \tag{4-177}$$

因为 $z\geqslant 0$ 时 $P(Z\leqslant z|Y=1)=1$，所以第一项是一个阶跃函数。对于第二项：

$$\begin{aligned} P(Z\leqslant z|Y=1) &= P(XY\leqslant z|Y=1) \\ &= P(X\leqslant z) = \lambda\exp(-\lambda z)u(z) \end{aligned} \tag{4-178}$$

因此，概率密度函数是混合的，如下式所示：

$$f_Z(z) = q\delta(z) + p\lambda\exp(-\lambda z) I_{[0,\infty)}(z) \tag{4-179}$$

利用式(4-167)得到：

$$f_Z(z) = \lambda\int_0^{\infty} (1/y)[p\delta(y-1)+q\delta(y)]\exp(-\lambda z/y)\mathrm{d}y$$

$$=p\lambda\exp(-\lambda z)I_{[0,\infty)}(z)+q\int_0^{\infty}(\lambda/y)\delta(y)\exp(-\lambda z/y)\mathrm{d}y \tag{4-180}$$

其中积分区间由 Y 的取值范围确定，它也是 Z 的取值范围。当 $z>0$ 时，因为指数衰减为 0 的速度比 $y\to 0$ 时 $1/y$ 趋近于无穷大的速度要快(注意指数里的参数还包括 $1/y$)，所以最后一个积分为 0。可以看出最后一个积分为狄拉克 δ 函数 $\delta(z)$(见习题 4-31)，再次得出式(4-179)中的混合概率密度函数。◀

4.9.4 商(比值)：$Z=X/Y$

两个随机变量商的累积分布函数如下：

$$F_Z(z)=P(X/Y\leqslant z)=\iint_{x/y\leqslant z}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y,$$

$$=\int_{-\infty}^{0}\int_{zy}^{\infty}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y+\int_0^{\infty}\int_{-\infty}^{zy}f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{4-181}$$

对 z 微分，并假设随机变量独立，得出：

$$f_Z(z)=-\int_{-\infty}^{0}yf_{X,Y}(zy,y)\mathrm{d}y+\int_0^{\infty}yf_{X,Y}(zy,y)\mathrm{d}y$$

$$=\int_{-\infty}^{\infty}|y|f_X(zy)f_Y(y)\mathrm{d}y \tag{4-182}$$

这是所谓的比值分布(Marsaglia，2006)，也称为商分布。对于离散随机变量，概率质量函数为：

$$p_Z[z]=\sum_y\sum_{x=zy}p_Y[y]p_X[x] \tag{4-183}$$

必须注意，因为除法 X/Y，Y 只在非 0 值有概率质量。

例 4-22 假设 X 和 Y 是独立同分布的高斯随机变量，参数为 $\{\mu=0,\sigma\}$。$Z=X/Y$ 的概率密度函数是：

$$f_Z(z)=\frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}|y|\exp(-(zy)^2/2\sigma^2)\exp(-y^2/2\sigma^2)\mathrm{d}y$$

$$=\frac{1}{\pi\sigma^2}\int_0^{\infty}y\exp(-y^2(z^2+1)/2\sigma^2)\mathrm{d}y \tag{4-184}$$

代入 $v\triangleq y\sqrt{z^2+1}$ 和 $\mathrm{d}v=\mathrm{d}y\sqrt{z^2+1}$，得到：

$$f_Z(z)=\frac{1}{\pi\sigma^2(z^2+1)}\int_0^{\infty}v\exp(-v^2/2\sigma^2)\mathrm{d}v \tag{4-185}$$

最后，令 $w\triangleq v^2/2\sigma^2$，因此 $\mathrm{d}w=v\mathrm{d}v/\sigma^2$ 和

$$f_Z(z)=\frac{1}{\pi(z^2+1)}\int_0^{\infty}\exp(-w)\mathrm{d}w=\frac{1}{\pi(z^2+1)} \tag{4-186}$$

这是标准柯西概率密度函数。◀

例 4-23 假设例 4-22 中的 X 和 Y 均值都是零，但有不同的方差 σ_X^2 和 σ_Y^2。比值的概率密度函数为：

$$f_Z(z)=\frac{1}{2\pi\sigma_X\sigma_Y}\int_{-\infty}^{\infty}|y|\exp(-(zy)^2/2\sigma_X^2)\exp(-y^2/2\sigma_Y^2)\mathrm{d}y$$

$$=\frac{1}{\pi\sigma_X\sigma_Y}\int_0^{\infty}y\exp(-y^2[z^2/2\sigma_X^2+1/2\sigma_Y^2])\mathrm{d}y$$

$$=\frac{1}{\pi\sigma_X\sigma_Y}\int_0^{\infty}y\exp(-y^2/2\alpha^2)\mathrm{d}y \tag{4-187}$$

其中已经代入了

$$z^2/2\sigma_X^2+1/2\sigma_Y^2=1/2\alpha^2 \tag{4-188}$$

从附录 A 可得，积分值与半正态分布的均值成正比：

$$\int_0^\infty y\exp(-y^2/2\alpha^2)\mathrm{d}y = a\alpha\sqrt{2/\pi} = \alpha^2 \tag{4-189}$$

比例常数 $a=\sqrt{\pi\alpha^2/2}$。代入该结果，得出最终的结果：

$$f_Z(z) = \frac{\alpha^2}{\pi\sigma_X\sigma_Y} = \frac{\sigma_X\sigma_Y}{\pi(\sigma_X^2 + z\sigma_Y^2)} = \frac{\sigma_X/\sigma_Y}{\pi[(\sigma_X/\sigma_Y)^2 + z^2]} \tag{4-190}$$

这是另一个柯西概率密度函数。 ◀

例 4-24 在前面的例子中，比值里的两个随机变量具有相同的区间。接下来考虑两个区间不同的离散随机变量的比值。设 X 服从参数为 p 的几何分布，Y 服从输出为 $\{-1, 1\}$ 的对称伯努利分布，这样我们不需要担心被零除。很显然，$Z=X/Y$ 的区间为 $\mathcal{Z}$。考虑以下的概率：

$$P(Z=z) = \begin{cases} P(X=0), & z=0 \\ P(X=z, Y=1), & z>0 \\ P(X=-z, Y=-1), & z<0 \end{cases} \tag{4-191}$$

由于 X 和 Y 独立，得到下面的概率质量函数

$$p_Z[z] = p\delta[z] + p^2q^zI_N[z] + pq^{-z+1}I_{\{\cdots,-1\}}[z] \tag{4-192}$$

接下来利用式(4-183)：

$$p_Z[z] = \sum_{y=-1,1}\sum_{x=zy,x\in\mathcal{Z}^+} pq^x p^{(1+y)/2} q^{(1-y)/2} \tag{4-193}$$

由于输出为±1，所以 p 和 q 的指数与标准伯努利分布不同(见附录 A)。对 y 求和后得到两项：

$$p_Z[z] = \sum_{x=z,x\in\mathcal{Z}^+} p^2q^x + \sum_{x=-z,x\in\mathcal{Z}^+} pq^{x+1} \tag{4-194}$$

从中我们发现，仅在 $z=0$，它们都对概率质量函数有贡献。注意到第二项中的 $x=-z$ 和 $x\in\mathcal{Z}^+$ 与 $z\in\mathcal{Z}^-$ 等价(包括零)，从而

$$p_Z[z] = p^2q^zI_{\mathcal{Z}^+}[z] + pq^{-z+1}I_{\mathcal{Z}^-}[z] \tag{4-195}$$

从每个分量中拆分出 $z=0$ 并相加，得出 $\delta[z]$ 的概率为 $p^2+pq=p$，因此上式与(4-192)是等价的。 ◀

4.10 随机变量簇的变换

本节将总结第 3 章讨论的一些参数分布的变换结果。利用前面的方法，仅讨论几个变换的推导过程。其余的在本章结尾的习题中讨论。

4.10.1 高斯变换

设 $\{X_1,\cdots,X_N\}$ 是独立同分布高斯随机变量，参数为 $\{\mu,\sigma\}$。表 4-3 中提供了几个变换的结果。

表 4-3 高斯随机变量的变换

变量	独立同分布 $\{x_n\}$ 的要求	随机变量 Y(参数)
$Y=\exp(X)$	无	log 正态 $\{\mu, \sigma\}$
$Y=\lvert X\rvert$	$\mu=0$	半正态 $\{\alpha=\sigma\}$
$Y=X_1/X_2$	$\mu=0$，$\sigma=1$	柯西 $\{c=0, \alpha=1\}$
$Y=X_1\pm X_2$	无	高斯 $\{0, 2\mu, 2\sigma\}$
$Y=\sqrt{X_1^2+X_2^2}$	$\mu=0$	瑞利 $\{\alpha=\sigma\}$
$Y=\sum_{n=1}^{N} X_n^2$	$\mu=0$，$\sigma=1$	卡方 $\{N\}$
$Y=\arctan(X_1/X_2)$	$\mu=0$，$\sigma=1$	均匀 $\{0, 2\pi\}$

例 4-25 ($Y=\exp(X)$)由于指数函数是单调的，采用累积分布函数方法得到：

$$F_Y(y)=P(\exp(X)\leqslant y)=P(X\leqslant \ln(y))$$
$$=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\ln(y)}\exp(-(x-\mu)^2/2\sigma^2)\,dx \tag{4-196}$$

微分后得到：

$$f_Y(y)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(\ln(y)-\mu)^2/2\sigma^2)\frac{d\ln(y)}{dy}$$
$$=\frac{1}{y\sqrt{2\pi\sigma^2}}\exp(-(\ln(y)-\mu)^2/2\sigma^2)I_{[0,\infty)}(y) \tag{4-197}$$

它是参数为$\{\mu,\sigma\}$的对数正态分布概率密度函数。利用概率密度函数(直接)方法同样可以很容易得出该结果。◀

例 4-26 ($Y=g(X_1/X_2)$)例4-22中利用式(4-182)中的通用公式导出了该概率密度函数。这里我们利用δ函数变换方法得出($\mu=0,\sigma=1$)

$$f_Y(y)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-(x_1^2+x_2^2)/2)\delta(x_1/x_2-y)\,dx_1\,dx_2$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}|x_2|\exp(-(x_2^2y^2+x_2^2)/2)\,dx_2$$
$$=\frac{1}{\pi}\int_0^{\infty}x_2\exp(-x_2^2(y^2+1)/2)\,dx_2 \tag{4-198}$$

其中

$$\delta(x_1/x_2-y)=|x_2|\delta(x_1-x_2y) \tag{4-199}$$

关于x_1积分时利用了筛选特性。注意到积分与式(4-189)类似，对变量进行适当的变换后，得到了以下的结果：

$$f_Y(y)=\frac{1}{\pi(1+y^2)} \tag{4-200}$$

它也是标准柯西分布。◀

4.10.2 指数变换

令$\{X_1,\cdots,X_N\}$是独立同分布的参数为λ的指数随机变量。表4-4给出了几个变换结果。

表 4-4　指数随机变量的变换

变换	独立同分布$\{X_n\}$的要求	随机变量 Y\{参数\}
$Y=aX$	$a>0$	指数 1$\{\lambda/a\}$
$Y=\exp(-\lambda X)$	无	均匀(0, 1]
$Y=\sqrt{2a^2X/\lambda}$	无	瑞利$\{a\}$
$Y=r\exp(X)$	$r\geqslant 1$	帕雷托$\{r,\ \alpha=\lambda\}$
$Y=\mu-\alpha\ln(\exp(-X)/(1-\exp(-X)))$	$\lambda=1$	Logistic$\{\mu,\ \alpha\}$
$Y=X^{1/r}$	$\lambda\to 1/\lambda^r$	韦伯$\{r,\ \alpha=1/\lambda\}$
$Y=X_1-X_2$	无	拉普拉斯$\{\mu=0,\ \alpha=1/\lambda\}$
$Y=\min(X_1,\ X_2)$	无	指数$\{2\lambda\}$
$Y=\sum_{n=1}^{N}X_n$	无	伽马$\{r=N,\ \alpha=\lambda\}$

例 4-27 ($Y=X_1-X_2$)。采用δ函数法：

$$f_Y(y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\lambda^2\exp(-\lambda(x_1+x_2))\delta(x_1-x_2-y)u(x_1)u(x_2)\,dx_1\,dx_2$$

$$=\lambda^2\int_{-\infty}^{\infty}\exp(-\lambda(2x_2+y))u(x_2+y)u(x_2)\mathrm{d}x_2 \tag{4-201}$$

两个阶跃函数都只影响积分下限：

$$f_Y(y)=\lambda^2\exp(-\lambda y)\int_{\min(0,-y)}^{\infty}\exp(-2\lambda x_2)\mathrm{d}x_2$$
$$=(\lambda/2)\exp(-\lambda y)\exp(-2\lambda\min(0,-y)) \tag{4-202}$$

考虑 y 的符号：

$$f_Y(y)=\begin{cases}(\lambda/2)\exp(-\lambda y), & y\geqslant 0\\(\lambda/2)\exp(-\lambda y+2\lambda y), & y<0\end{cases} \tag{4-203}$$

得到：

$$f_Y(y)=(\lambda/2)\exp(-\lambda|y|) \tag{4-204}$$

它是均值为 0，比例参数 $\alpha=1/\lambda$ 的拉普拉斯分布的概率密度函数。◀

例 4-28 ($Y=\exp(-\lambda X)$)。利用概率密度函数(直接)法：

$$f_Y(y)=\frac{f_X(x_n)}{|g'(x_n)|}\bigg|_{x_n=g^{-1}(y)}$$
$$=\frac{\lambda\exp(-\lambda x)u(x)}{|\lambda\exp(-\lambda x)|}\bigg|_{x_n=g^{-1}(y)}=u(x)|_{x_n=g^{-1}(y)} \tag{4-205}$$

其中代入了 $g'(x)=-\lambda\exp(-\lambda x)$。由于 $g(x)$ 是单调的，因此只有一个 $x_n=x$，并且由于 $x\in[0,+\infty)$，$y\in(0,1]$，代入 $x=-\ln(y)/\lambda$ 得到：

$$f_Y(y)=u(-\ln(y)/\lambda)=I_{(0,1]}(y) \tag{4-206}$$

这是 (0,1] 上均匀分布的随机变量的概率密度函数。由于 $u(-\ln(y)/\lambda)=1$，当 $-\ln(y)/\lambda>0\Rightarrow\ln(y)<0$，意味着 $y\in(0,1]$，从而得到最后一个表达式。◀

4.10.3 卡方变换

设 $\{X_1,X_2\}$ 是独立的自由度分别为 $\{m,n\}$ 的卡方随机变量。表 4-5 中给出了几个变换的结果。

表 4-5 卡方随机变量的变换

变换	独立 $\{X_p\}$ 的要求	随机变量 Y {参数}
$Y=aX$	$a>0$	伽马 $\{r=n/2,\ \lambda=1/2a\}$
$Y=\dfrac{X_1/m}{X_2/n}$	无	$F\{m,n\}$
$Y=X_1/\sqrt{X_2/r}$	X_1，标准高斯	学生氏 t$\{r\}$
$Y=X_1+X_2$	无	卡方 $\{N=2n\}$
$Y=X_1/(X_1+X_2)$	无	贝塔 $(\alpha=n/2,\beta=n/2)$

例 4-29 ($Y=(X_1/m)/(X_2/n)$) 由 δ 函数方法得到：

$$f_Y(y)=\frac{1}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1^{m/2-1}x_2^{n/2-1}$$
$$\cdot\exp(-(x_1+x_2)/2)\delta([(x_1/m)/(x_2/n)]-y)u(x_1)u(x_2)\mathrm{d}x_1\mathrm{d}x_2 \tag{4-207}$$

代入：

$$\delta(nx_1/mx_2-y)=(m/n)|x_2|\delta(x_1-(m/n)x_2y) \tag{4-208}$$

得到：

$$f_Y(y)=\frac{1}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)}\int_{-\infty}^{\infty}(m/n)|x_2|((m/n)x_2y)^{m/2-1}x_2^{n/2-1}$$
$$\cdot\exp(-((m/n)x_2y+x_2)/2)u((m/n)x_2y)u(x_2)\mathrm{d}x_2$$

$$= \frac{(m/n)^{m/2}}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)} y^{m/2-1} \int_0^\infty x_2^{(m+n)/2-1} \cdot \exp(-((m/n)y+1)x_2/2)\mathrm{d}x_2 \tag{4-209}$$

如果将积分值乘以$[(m/n)y+1]/2$，则乘积是指数概率密度函数的非中心矩(见附录A)：

$$\frac{(m/n)y+1}{2}\int_0^\infty x_2^{(m+n)/2-1}\exp(-((m/n)y+1)x_2/2)\mathrm{d}x_2 = \frac{\Gamma((m+n)/2-1)}{([(m/n)y+1]/2)^{(m+n)/2-1}} \tag{4-210}$$

因此：

$$f_Y(y) = \frac{(m/n)^{m/2}}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)} \frac{\Gamma((m+n)/2)}{([(m/n)y+1]/2)^{(m+n)/2-1}} \frac{y^{m/2-1}}{((m/n)y+1)/2} \tag{4-211}$$

整理并代入贝塔函数：

$$B(m/2,n/2) \triangleq \frac{\Gamma(m/2+n/2)}{\Gamma(m/2)\Gamma(n/2)} \tag{4-212}$$

得到：

$$f_Y(y) = \frac{(m/n)^{m/2} y^{m/2-1}}{B(m/2,n/2)((m/n)y+1)^{(m+n)/2}} \tag{4-213}$$

这是参数为$\{m,n\}$的F分布。 ◀

4.11　随机向量的变换

前面介绍的一个随机变量变换到另一个随机变量的方法可以扩展到多个随机变量。我们不会详细推导所有的结果，而是在推导连续随机变量的变换后提出主要思路，然后总结包括离散随机变量在内的各种类型变换的结果。

考虑随机向量$\boldsymbol{X} \triangleq [X_1,\cdots X_N]^{\mathrm{T}}$，以及根据$Y_n = g_n(\boldsymbol{X})$得到的变换后的向量$\boldsymbol{Y} = [Y_1,\cdots Y_N]^{\mathrm{T}}$，其中$\{g_n(\cdot)\}$是已知函数(可能是非线性的)。这些函数可能严格递增、严格递减、或者非单调，也有可能包含平坦区域或者在多个不同的区间包含这些特点的组合。该函数也可能将$\boldsymbol{X}$的一个子集映射到$\boldsymbol{Y}$。首先考虑N个函数是严格递增或严格递减的。对于这种情况，逆象$\{g_n^{-1}(\boldsymbol{Y})\}$是有明确定义的。

- 连续的随机向量。连续随机向量之间的变换是式(4-106)中结果的直接扩展，利用直接方法将一个随机变量映射到另一个随机变量。$\boldsymbol{Y}$的联合概率密度函数是：

  $$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |J(\boldsymbol{x})|^{-1} f_{\boldsymbol{X}}(\boldsymbol{x})|_{x_n=g_n^{-1}(y)} \tag{4-214}$$

  其中$n=1,\cdots,N$，$J(\boldsymbol{x})$是雅可比行列式，计算方法如下

  $$J(\boldsymbol{x}) \triangleq \det[\boldsymbol{J}(\boldsymbol{x})] = \det\begin{bmatrix} \partial y_1/\partial x_1 & \cdots & \partial y_1/\partial x_N \\ \vdots & \ddots & \vdots \\ \partial y_N/\partial x_1 & \cdots & \partial y_N/\partial x_N \end{bmatrix} \tag{4-215}$$

  其中$\boldsymbol{J}(\boldsymbol{x})$是雅可比矩阵(见附录G)。$\boldsymbol{J}(\boldsymbol{x})$中的变量为独立变量，因此矩阵的元素是$\boldsymbol{x}$的函数。需要注意的是(4-214)使用的是$J(\boldsymbol{x})$的绝对值，不要与(4-215)中可能是正或负的行列式混淆。通过对每个函数$g_n(\boldsymbol{X})$求映射到同一y_n的不同$\boldsymbol{X}$的分量，可以将前面学习的一个随机变量变换的结论推广到包括多个非单调函数的情况。此外，任意雅可比偏导数为0的区域将导致$f_Y(\boldsymbol{y})$包含离散部分，从而产生混合概率密度函数。

对于某些映射，更容易计算逆变换的雅可比矩阵，如下所示：

$$J(\boldsymbol{y}) \triangleq \det\begin{bmatrix} \partial x_1/\partial y_1 & \cdots & \partial x_1/\partial y_N \\ \vdots & \ddots & \vdots \\ \partial x_N/\partial y_1 & \cdots & \partial x_N/\partial y_N \end{bmatrix} \tag{4-216}$$

因此利用式(4-217)代替式(4-214)：

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |J(\boldsymbol{y})| f_{\boldsymbol{X}}(\boldsymbol{x})|_{x_n=g_n^{-1}(y)} \tag{4-217}$$

这个表达式中的雅可比行列式是 $\mathbf{y}$ 的函数，因此，$x_n=g_n^{-1}(\mathbf{y})$ 只用于 $f_{\mathbf{X}}(\mathbf{x})$。

例 4-30 设 X_1 和 X_2 是独立的标准高斯随机变量，考虑以下变换：

$$Y_1=\sqrt{X_1^2+X_2^2},Y_2=\arctan(X_1/X_2) \tag{4-218}$$

显然 Y_1 取值范围为 $\mathcal{R}^+$，Y_2 为 $[0,2\pi]$。为得到联合分布 $f_{\mathbf{Y}}(\mathbf{y})$，我们需要得到(ⅰ)雅可比行列式和(ⅱ)逆变换。对于(ⅰ)：

$$\begin{aligned}J(x_1,x_2)&=\det\begin{bmatrix}x_1/\sqrt{x_1^2+x_2^2} & x_2/\sqrt{x_1^2+x_2^2}\\ x_2/(x_1^2+x_2^2) & -x_1/(x_1^2+x_2^2)\end{bmatrix}\\&=-1/\sqrt{x_1^2+x_2^2}\end{aligned} \tag{4-219}$$

对于(ⅱ)，我们按照下式重新整理变换：

$$Y_1^2=X_1^2+X_2^2,X_1^2=X_2^2\tan^2(Y_2) \tag{4-220}$$

利用上面两个式子求解 X_1 和 X_2，得到

$$X_1=Y_1\sin(Y_2),X_2=Y_1\cos(Y_2) \tag{4-221}$$

将这些等式代入(4-214)，得到

$$\begin{aligned}f_{\mathbf{Y}}(\mathbf{y})&=(x_1^2+x_2^2)^{-1/2}(1/2\pi)\exp(-(1/2)[x_1^2+x_2^2])\,|_{x_n=g_n^{-1}(\mathbf{y})}\\&=|y_1|(1/2\pi)\exp(-(1/2)(y_1^2\sin^2(y_2)+y_1^2\cos^2(y_2)))\\&=(1/2\pi)I_{[0,2\pi]}(y_2)|y_1|\exp(-y_1^2/2)I_{\mathcal{R}^+}(y_1)\end{aligned} \tag{4-222}$$

从中我们可以看出，Y_1 和 Y_2 独立，分别服从瑞利分布和均匀分布。为完整性，请注意另一个雅可比行列式为：

$$J(y_1,y_2)=\det\begin{pmatrix}\sin(y_2) & y_1\cos(y_2)\\ \cos(y_2) & -y_1\sin(y_2)\end{pmatrix}=-y_1 \tag{4-223}$$

其大小是(4-222)第二行的第一个数。◀

- **离散随机向量** 对于离散随机向量之间的转换，我们需要找到变换后所有可能的输出，然后确定 $\mathbf{X}$ 的每个分量的概率质量是怎样作用于 $\mathbf{Y}$ 的每个分量的概率质量的。下面的例子说明了该方法。

例 4-31 设 $\mathbf{X}$ 有 $N=2$ 个独立变量，分别是参数为 α_1 和 α_2 的泊松随机变量。考虑下列变换：

$$Y_1=X_1+X_2,\quad Y_2=X_1^2 \tag{4-224}$$

因此 Y_1 和 X_1、X_2 具有同样范围，但 Y_2 只包括具有整数平方根的整数。对于随机变量 Y_2，

$$p_{Y_2}[y_2]=P(X_1^2=y_2)=P(X_1=\sqrt{y_2})=\frac{\alpha_1^{\sqrt{y_2}}}{\sqrt{y_2}\,!}\exp(-\alpha_1) \tag{4-225}$$

其中 $y_2\in\{0,1,4,9,16,\cdots\}$。它不是标准参数分布。对于随机变量 Y_1

$$\begin{aligned}p_{Y_1}[y_1]&=\exp(-(\alpha_1+\alpha_2))\sum_{x=-\infty}^{\infty}\frac{\alpha_1^x}{x!}\frac{\alpha_2^{y_1-x}}{(y_1-x)!}I_{\mathcal{Z}^+}[x]I_{\mathcal{Z}^+}[y_1-x]\\&=\exp(-(\alpha_1+\alpha_2))\sum_{x=0}^{y_1}\frac{\alpha_1^x\alpha_2^{y_1-x}}{x!(y_1-x)!}I_{\mathcal{Z}^+}[y_1],\end{aligned} \tag{4-226}$$

第一个方程中的第二个指示函数导致求和的上限是 y_1，因为 $y_1-x\in\mathcal{Z}^+$ 意味着 x 不能超过 y_1(否则，y_1-x 为负)。采用二项式公式，如果我们乘以和除以 $y_1!$，改写得到的最后结果为：

$$\begin{aligned}p_{Y_1}[y_1]&=(1/y!)\exp(-(\alpha_1+\alpha_2))\sum_{x=0}^{y_1}\frac{\alpha_1^x\alpha_2^{y_1-x}y_1!}{x!(y_1-x)!}I_{\mathcal{Z}^+}[y_1]\\&=(1/y_1!)\exp(-(\alpha_1+\alpha_2))\sum_{x=0}^{y_1}\binom{y_1}{x}\alpha_1^x\alpha_2^{y_1-x}I_{\mathcal{Z}^+}[y_1]\\&=(1/y!)\exp(-(\alpha_1+\alpha_2))(\alpha_1+\alpha_2)^{y_1}I_{\mathcal{Z}^+}[y_1]\end{aligned} \tag{4-227}$$

这是参数为 $\alpha_1+\alpha_2$ 的泊松分布。◀

4.12　样本均值 $\overline{X}$ 和样本方差 S^2

样本均值是一些随机变量的最重要的函数之一，又称为算术平均值。(几何平均是 $(X_1,\cdots,X_N)^{1/N}$)。考虑一个重复试验，如连续掷硬币，观察到 N 个输出，对应 N 个随机变量 $\{X_1,\cdots,X_N\}$。这种重复试验可被看作是第 6 章学习过的一个离散时间随机序列。本节我们只需要考虑 N 个随机变量，而不考虑它们如何生成和是否相互依赖。样本均值通常用来衡量输出的均值：它是均值的一个估计量。

定义(样本均值)　随机变量 $\{X_1,\cdots,X_N\}$ 的**样本均值**为：

$$\overline{X}\triangleq\frac{1}{N}\sum_{n=1}^{N}X_n \tag{4-228}$$

它本身就是一个随机变量。

在随后的章节中，我们将讨论 $\overline{X}$ 的一些有意思的性质。在这里，尽管样本均值是独立同分布$\{X_n\}$的 μ_X 的无偏估计，但由于它对异常值敏感，所以不是鲁棒的。尽管大的输出(正或负)不太可能发生，但也有机会发生，导致样本均值小于或大于实际均值。要提高估计器的性能，一种方法是增大 N，从而将异常值的影响减小到最小。或者可以使用其他均值估计器(或概率密度函数的“中心”)，如在第 5 章中讨论的中位数。

定义(样本方差)　随机变量 $\{X_1,\cdots,X_N\}$ 的**样本方差**为：

$$S^2\triangleq\frac{1}{N-1}\sum_{n=1}^{N}(X_n-\overline{X})^2 \tag{4-229}$$

它本身就是一个随机变量。

其中包括一个因子 $1/(N-1)$(而不是 $1/N$)，因此 S^2 是独立同分布$\{X_n\}$的 σ_X^2 的无偏估计。

例 4-32　设$\{X_n\}$在$[-1, 1]$上均匀分布。当 $N=2$ 时，$\overline{X}$ 的概率密度函数可以通过对两个相同的矩形概率密度函数卷积获得，得到一个三角形概率密度函数，如图 4-34a 所示。增大 N，使矩形概率密度函数与前面的卷积结果再次进行卷积。增大 N 以后得到连续概率密度函数，如图 4-34a 所示。观察发现样本均值的概率密度函数开始类似于一个高斯分布。这种现象是前面讨论过的(在第 7 章将再次看到)CLT 的一个特征，它的条件是，当 $N\to\infty$ 时，独立随机变量之和趋近于高斯概率密度函数。另一个样本均值的概率密度函数的例子如图 4-34b 所示，其中样本服从参数 $\lambda=1$ 的指数分布。请注意，即使指数概率密度函数不对称(像均匀和高斯概率密度函数那样)，样本均值的概率密度函数的形状也迅速趋近高斯概率密度函数。

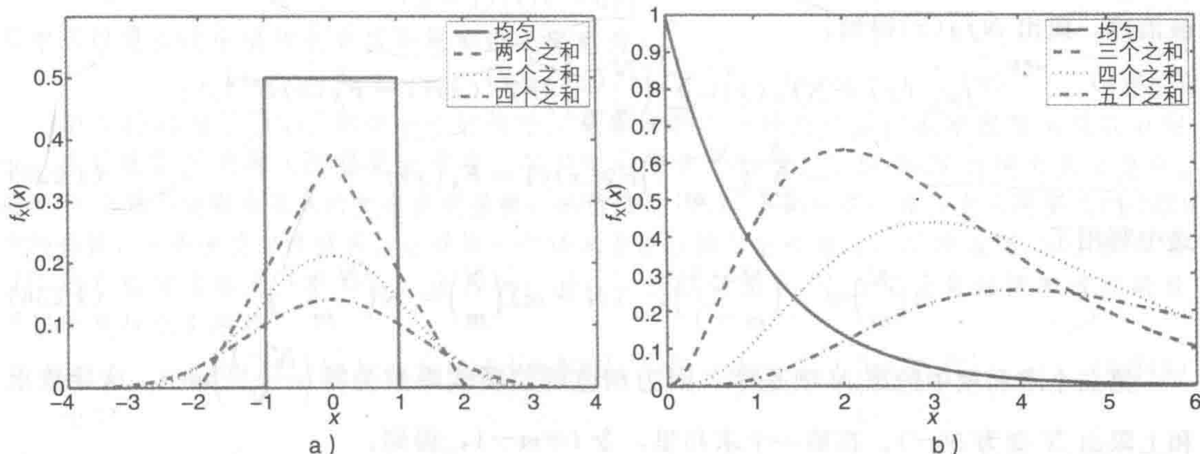


图 4-34　N 增大时，样本均值的概率密度函数。a) $\{X_n\}$是独立同分布的 $[-1,1]$ 上的均匀分布；b) $\{X_n\}$是独立同分布的 $\lambda=1$ 的指数分布。为方便观察，概率密度函数乘以了一定的系数 ◀

4.13 最小值、最大值和顺序统计量

考虑 N 个独立同分布的随机变量 $\{X_1,\cdots,X_N\}$，其概率密度函数为 $f_X(x)$。本节的目的是推导出最小 $X_{(1)} \triangleq \min(X_1,\cdots,X_N)$ 的概率密度函数（称为一阶统计量），最大 $X_{(N)} \triangleq \max(X_1,\cdots,X_N)$（称为第 N 阶统计量），以及整个顺序统计量。

定义（顺序统计量） **顺序统计量**是将随机变量 $\{X_1,\cdots,X_N\}$ 按升序排列：$\{X_{(1)},\cdots,X_{(N)}\}$。

由于随机变量是随机的，任何一个都有取值范围，因此对随机变量进行按大小排序时可能会有问题。当然，一旦我们有一个实际的样本（采集输出）$\{x_1,\cdots,x_N\}$ 可以对它们按大小排序。然而，排序好的随机变量 $\{X_{(1)},\cdots,X_{(N)}\}$ 可从概率观点描述，如我们现在使用的累积分布函数变换方法。

定理 4-4（顺序统计量） 第 n 阶统计量 $X_{(n)}$ 的概率密度函数为：

$$f_{X(n)}(x)=\frac{N!}{(n-1)!(N-n)!}F_X^{n-1}(x)[1-F_X(x)]^{N-n}f_X(x) \tag{4-230}$$

其中，$f_X(x)$ 和 $F_X(x)$ 分别是独立同分布 $\{X_1,\cdots,X_N\}$ 的概率密度函数和累积分布函数，$n\in\{1,\cdots,N\}$。

证明：第 n 阶统计量的累积分布函数为：

$$\begin{aligned}F_{X_{(n)}}(x)&=P(X_{(n)}\leqslant x)=P(\text{最少 } n \text{ 个 } X_m\leqslant x)\\&=\sum_{m=n}^{N}\binom{N}{m}P(X\leqslant x)^m(1-P(X\leqslant x))^{N-m}\\&=\sum_{m=n}^{N}\binom{N}{m}F_X^m(x)(1-F_X(x))^{N-m}\end{aligned} \tag{4-231}$$

关键的步骤是第一行：因为我们感兴趣的是排序后的样本，$X_{(n)}\leqslant x$ 要求至少有 n 个随机变量 $\leqslant x$。表达式里允许随机变量数目超过 n，因为我们要保证第 n 阶统计量 $\leqslant x$。求和式中的表达式是序列的数目，其中有 n 个或更多 $N\{X_m\}\leqslant x$。该结果对应于成功概率为 $p=F_X(x)=P(X\leqslant x)$ 和失败概率为 $q=1-F_X(x)=P(X>x)$ 的二项概率密度函数（注意，$0\leqslant F_X(x)\leqslant 1$ 是二项分布概率密度函数中的 p 要求的）。只要 n 个或更多的随机变量 $\leqslant x$，则概率非零。关于 x 进行微分并使用链式法则，得到概率密度函数：

$$f_{X_{(n)}}(x)=\sum_{m=n}^{N}\binom{N}{m}f_X(x)[mF_X^{m-1}(x)(1-F_X(x))^{N-m}-(N-m)F_X^m(x)(1-F_X(x))^{N-m-1}] \tag{4-232}$$

虽然不是很明显，这个式子仍可以简化。观察到被减项只有一项不同，其他项都可以被消掉。提出 $Nf_X(x)$ 得到：

$$\begin{aligned}f_{X_{(n)}}(x)=Nf_X(x)\Big[&\sum_{m=n}^{N}\binom{N-1}{m-1}F_X^{m-1}(x)(1-F_X(x))^{N-m}\\&-\sum_{m=n}^{N}\binom{N-1}{m}F_X^m(x)(1-F_X(x))^{N-m-1}\Big]\end{aligned} \tag{4-233}$$

式中利用了

$$m\binom{N}{m}=N\binom{N-1}{m-1},\quad (N-m)\binom{N}{m}=N\binom{N-1}{m} \tag{4-234}$$

第二个求和项中的第 N 项是零，因为相应的二项式系数为零 $\binom{N-1}{N}\triangleq 0$。这导致求和上限由 N 变为 $N-1$。在第一个求和里，令 $l=m-1$，得到：

$$f_{X_{(n)}}(x)=Nf_X(x)\Big[\sum_{l=n-1}^{N-1}\binom{N-1}{l}F_X^l(x)(1-F_X(x))^{N-1-l}$$

$$- \sum_{m=n}^{N-1} \binom{N-1}{m} F_X^m(x)(1-F_X(x))^{N-m-1} \Big] \tag{4-235}$$

经过减法，第一个求和里只有第 $n-1$ 项保留下来

$$f_{X_{(n)}}(x) = Nf_X(x)\binom{N-1}{n-1}F_X^{n-1}(x)(1-F_X(x))^{N-n} \tag{4-236}$$

它与式(4-230)相同。

定理 4-5(一阶统计量：最小)　在 $\{X_1,\cdots,X_N\}$ 独立同分布条件下，$X_{(1)} \triangleq \min(X_1,\cdots,X_N)$ 的概率密度函数是：

$$f_{X_{(1)}}(x) = N[1-F_X(x)]^{N-1} f_X(x) \tag{4-237}$$

累积分布函数是：

$$F_{X_{(1)}}(x) = 1-[1-F_X(x)]^N \tag{4-238}$$

证明：该概率密度函数是定理 4-4 的一种特殊情况，代入 $n=1$ 可以很容易证明。累积分布函数推导如下：

$$\begin{aligned} F_{X_{(1)}}(x) &= 1-P(X_1 > x,\cdots,X_N > x) \\ &= 1-[1-F_1(x)]\cdots[1-F_N(x)] = 1-[1-F_X(x)]^N \end{aligned} \tag{4-239}$$

这也可以通过将 $n=1$ 代入(4-231)得到，它是除了 $m=0$ 以外所有二项分布的总和。

定理 4-6(第 N 阶统计量：最大)　在 $\{X_1,\cdots,X_N\}$ 独立同分布条件下，$X_{(N)} \triangleq \max(X_1,\cdots,X_N)$ 的概率密度函数是：

$$f_{X_{(N)}}(x) = NF_X^{N-1}(x)f_X(x) \tag{4-240}$$

累积分布函数是：

$$F_{X_{(N)}}(x) = F_X^N(x) \tag{4-241}$$

证明：该概率密度函数是定理 4-4 的一种特殊情况，代入 $n=N$ 可以很容易证明。累积分布函数推导如下：

$$F_{X_{(N)}}(x) = P(X_1 \leqslant x,\cdots,X_N \leqslant x) = F_1(x)\cdots F_N(x) = F_X^N(x) \tag{4-242}$$

也可以通过将 $n=N$ 代入(4-231)得到。

例 4-33　令 $\{X_1,\cdots,X_N\}$ 是[0,1]上独立同均匀分布的样本，其概率密度函数和累积分布函数为：

$$f_X(x) = I_{[0,1]}(x), \quad F_X(x) = xI_{[0,1]}(x) \tag{4-243}$$

利用定理 4-4 得出第 n 阶统计量的概率密度函数为：

$$f_{X_{(n)}}(x) = \frac{N!}{(n-1)!(N-n)!}x^{n-1}(1-x)^{N-n}I_{[0,1]}(x) \tag{4-244}$$

从中可以得出最小值和最大值的概率密度函数为：

$$f_{X_{(1)}}(x) = N(1-x)^{N-1}I_{[0,1]}(x), \quad f_{X_{(N)}}(x) = Nx^{N-1}I_{[0,1]}(x) \tag{4-245}$$

图 4-35 给出了[0,1]均匀分布的例子。观察发现，一阶统计量的概率密度函数在 0 附近，并且随着 N 的增大变得更加集中。N 阶统计量在 1 附近，也随着 N 的增大更加集中。增大 N 当然不会影响基本的概率密度函数，其仍然是[0,1]上的均匀分布。它只改变式(4-242)中的指数，从而使得样本增多，这样第一个样本靠近 0 的可能性越大，同样地第 N 个样本靠近 1 的可能性也越大。事实上，当 $N\to\infty$ 时，一阶统计量和 N 阶统计量的概率密度函数近似为狄拉克 δ 函数：

$$\lim_{N\to\infty} f_{X_{(N)}}(x) = f_X(x) \lim_{N\to\infty} NF_X^{N-1}(x) = I_{[0,1]}(x) \lim_{N\to\infty} Nx^{N-1} \tag{4-246}$$

由于

$$\int_{-\infty}^{\infty} Nx^{N-1}I_{[0,1]}(x)\mathrm{d}x = 1 \tag{4-247}$$

对于任意 $N>0$，上述极限基本上是狄拉克 δ 函数的另一个定义(见附录 B)：$N\to\infty$ 时，在

$x=1$ 处函数的高值趋近于无穷大，宽度接近为零，并且它的面积固定为 1。一阶统计量在 $x=0$ 处有同样的现象。图中还有 $n=2$ 和 40 阶统计量，分别对应 $N=$ 为 5 和 50。

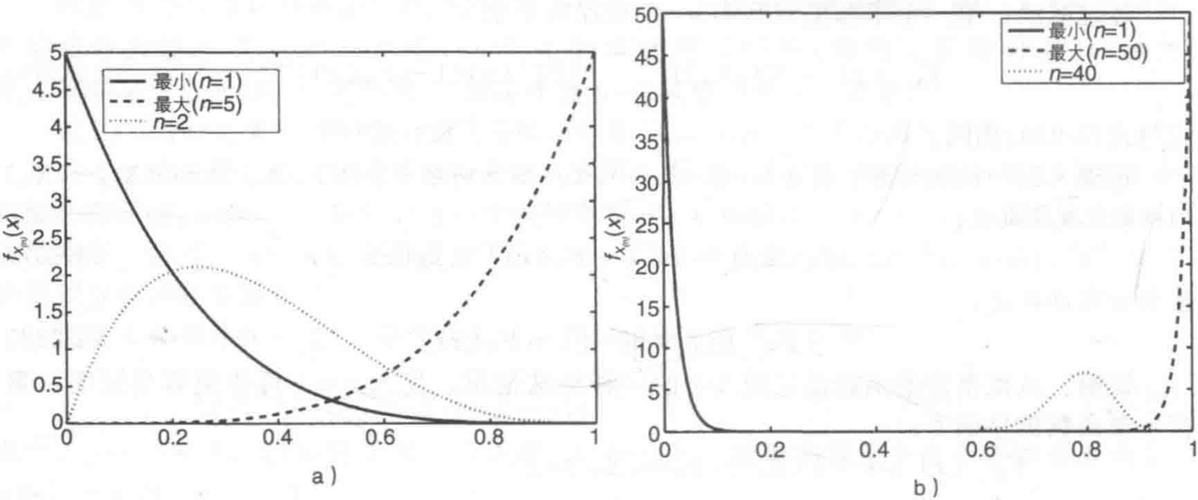


图 4-35 例 4-33 中 [0,1] 均匀分布随机变量的顺序统计量的概率密度函数。a) 样本数 $N=5$；b) 样本数 $N=50$ ◀

例 4-34 考虑标准高斯随机变量的顺序统计量。最小和最大的概率密度函数为：

$$f_{X_{(1)}}(x)=\frac{NQ^{N-1}(x)}{\sqrt{2\pi}}\exp(-x^2/2) \tag{4-248}$$

$$f_{X_{(N)}}(x)=\frac{N[1-Q(x)]^{N-1}}{\sqrt{2\pi}}\exp(-x^2/2) \tag{4-249}$$

式中，$Q(x)$是 Q 函数。概率密度函数如图 4-36 所示。由于概率密度函数不具有有限范围，样本不像均匀分布那样在最大值或最小值附近"叠加"。相反，随着 N 的增加，概率密度函数不断向左和向右移动，这是很有意义的，因为样本越多，越有可能出现一个非常小和一个非常大的样本。

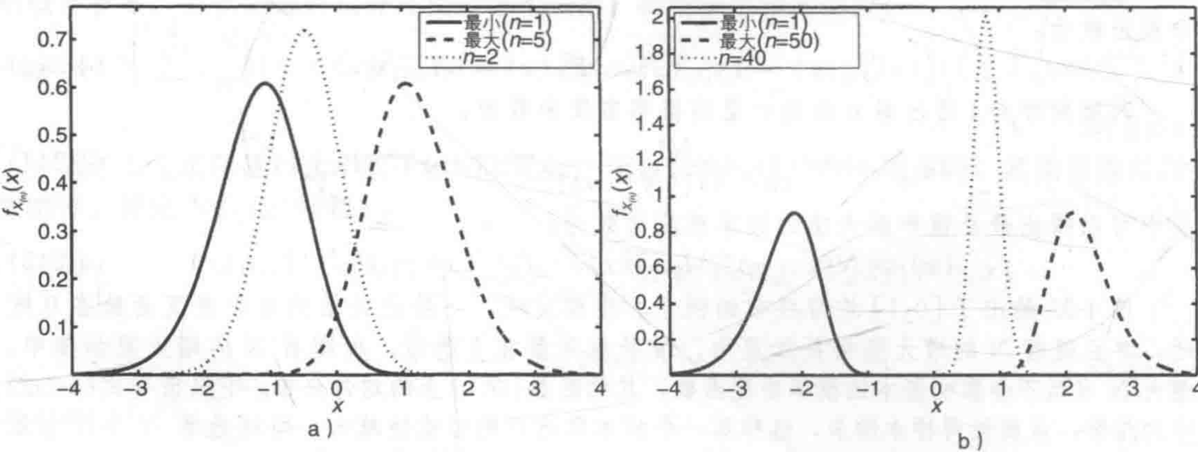


图 4-36 例 4-34 中标准高斯分布的顺序统计量的概率密度函数。a) 样本数 $N=5$；b) 样本数 $N=50$ ◀

4.14 混合

用$\{f_{X_n}(x_n)\}$表示随机变量$\{X_n\}$, $n=1,\cdots,N$ 的概率密度函数。

定义(概率混合) 如果随机变量 Y 可以表示为下列加权和，那么它是一个**概率混合**：

$$f_Y(y)=\sum_{n=1}^{N}a_n f_{X_n}(y) \tag{4-250}$$

其中，$a_n\in[0,1]$ 并且 $\sum_{n=1}^{N}a_n=1$，这也称为有限混合。

混合不像样本均值那样是随机变量之和。式(4-250)中的和是一个凸组合，因为系数非负且总和为 1。这两个属性确保 $f_Y(y)$是一个有效的概率密度函数，即 $f_Y(y)\geqslant 0$ 并且：

$$\begin{aligned}\int_{-\infty}^{\infty}f_Y(y)\mathrm{d}y&=\int_{-\infty}^{\infty}\sum_{n=1}^{N}a_n f_{X_n}(y)\mathrm{d}y\\&=\sum_{n=1}^{N}a_n\int_{-\infty}^{\infty}f_{X_n}(y)\mathrm{d}y=\sum_{n=1}^{N}a_n=1\end{aligned} \tag{4-251}$$

需要注意的是，式(4-250)中可以使用不同的概率密度函数。如果概率密度函数来自同一概率簇，只是概率分布参数不同，概率混合可称为参数混合。

定义(参数混合(离散))　当随机变量 Y 可以写成以下加权和时，它是一个参数混合：

$$f_Y(y)=\sum_{n=1}^{N}a_n f_X(y;\theta_n) \tag{4-252}$$

其中，$a_n\in[0,1]$ 并且 $\sum_{n=1}^{N}a_n=1$，θ_n 是随机变量 X 的一个参数。

上述和只针对不同的参数值，求和项中只有一种类型(“簇”)的概率密度函数。参数连续时，也可以定义相应的混合。

定义(参数混合(连续))　当随机变量 Y 可以写成下列积分时，它是一个连续混合：

$$f_Y(y)=\int_{-\infty}^{\infty}a(\theta)f_X(y;\theta)\mathrm{d}\theta \tag{4-253}$$

其中，$a(\theta)\geqslant 0$ 并且 $\int_{-\infty}^{\infty}a(\theta)\mathrm{d}\theta=1$，它也称为无限混合。

$a(\theta)$虽然满足概率密度函数的性能要求，但它不必是一个概率密度函数。如果 $a(\theta)$恰好为一个有效的概率密度函数，则 θ 可以看作是一个随机变量，我们可以利用 $f_\Theta(\theta)$明确定义它。

例 4-35　负二项随机变量 Y 是连续混合的一个例子(虽然 Y 本身是离散的)。设 X 是一个参数为 α 的泊松分布随机变量，并且假设 α 是参数为 $\{\lambda=p/q,r\}(q\triangleq 1-p)$ 的伽马分布随机变量，由式(4-253)：

$$\begin{aligned}f_Y(y)&=\int_{-\infty}^{\infty}f_X(y;\alpha)f_\alpha(\alpha)\mathrm{d}\alpha\\&=\int_{-\infty}^{\infty}\sum_{n=0}^{\infty}\exp(-\alpha)\frac{\alpha^n}{n!}\delta(y-n)\\&\quad\cdot\frac{1}{\Gamma(r)}\alpha^{r-1}(p/q)^r\exp(-(p/q)\alpha)u(\alpha)\mathrm{d}\alpha\end{aligned} \tag{4-254}$$

从例 3-62 中看出这个表达式可以重新表示为负二项概率密度函数的形式：

$$f_Y(y)=\sum_{n=0}^{\infty}\binom{n+r-1}{n}p^r q^n\delta(y-n) \tag{4-255}$$

其中参数为 $\{p,N=r\}$。◀

习题

随机变量的近似

4-1　证明当 $N\to\infty$时，参数为 N 的卡方随机变量服从参数为 $\{\mu=N,\sigma=\sqrt{2N}\}$ 的高斯分布。

4-2　证明参数为 r 的学生氏 t 分布，在 r 比较大时，近似服从标准高斯分布。

4-3　证明当 $N\to\infty$时，参数为$\{N,p\}$的负二项分布是参数 $\alpha=N(1-p)$的泊松分布。

4-4　证明当 $r\to\infty$时，参数为$\{\lambda,r\}$的伽马分布

是参数为 $\{\mu=r/\lambda,\sigma=\sqrt{r}/\lambda\}$ 的高斯分布。

联合、边缘和条件分布

4-5 (a)求下列联合概率密度函数的边缘概率密度函数：

$$f_{X,Y}(x,y)=y\exp(-x-y)I_{[0,\infty)}(x)I_{[0,\infty)}(y) \quad (4\text{-}256)$$

(b)求条件概率密度函数 $f_{Y|X}(y|x)$ 和 $f_{X|Y}(x|y)$。

4-6 对于下列的(a)和(b)重复习题 4-5。

(a)

$$f_{X,Y}(x,y)=(1/2)xyI_{[0,2]}(x)I_{[0,x]}(y) \quad (4\text{-}257)$$

(b)

$$f_{X,Y}(x,y)=(1/2)(3x+y)I_{[0,1]}(x)I_{[0,1]}(y) \quad (4\text{-}258)$$

4-7 顺序掷两个骰子。令 $X=$\{总和是偶数\}、$Y=$\{第一个骰子是偶数\}。判断 X 和 Y 是否独立。

4-8 设 X 是 $[0,a]$，$a>0$ 上的均匀分布，定义事件 $B=\{X>b\}$，其中 $0<b<a$。推导 $f_{X|B}(x|B)$ 的表达式。

4-9 关于式(4-80)中的二维指数概率密度函数，计算(a)条件概率密度函数 $f_{X_1|X_2}(x_1|x_2)$ (b)$P(X_1>X_2|X_2=2)$的表达式。

4-10 导出式(4-85)中二维指数随机变量累积分布函数的一个版本。

4-11 二维指数分布的另一个版本的累积分布函数为：

$$F_{X_1,X_2}(x_1,x_2)=[1-\exp(-x_1)-\exp(-x_2)+\exp(-\sqrt{x_1^2+x_2^2})]\cdot I_{[0,\infty)}(x_1)I_{[0,\infty)}(x_2) \quad (4\text{-}259)$$

计算(a)二维概率密度函数及(b)边缘概率密度函数的表达式。

4-12 设 X_1 和 X_2 是独立参数分别为 α_1 和 α_2 的泊松随机变量。利用习题 4-25 的结论，$Y=X_1+X_2$ 是参数为 $\alpha_1+\alpha_2$ 的泊松分布。证明当条件是 $Y=y$(一个固定的数字)时，X_1 的条件概率质量函数为二项分布。

4-13 考虑以下联合概率质量函数：

$$p_{X,Y}[x,y]=c\alpha^{x+y}I_{\{x,x+1,\cdots\}}[y]I_{\{0,1,\cdots\}}[x] \quad (4\text{-}260)$$

其中 $\alpha\in(0,1)$。(a)求 c，保证它是有效概率质量函数；(b)推导相应的两个边缘概率质量函数；(c)求条件概率质量函数 $p_{Y|X}[y|x]$。

4-14 求例 4-6 中互信息的表达式。

随机变量的变换

4-15 对于参数为 λ 的指数随机变量，证明变换 $Y=\sqrt{2\sigma^2X/\lambda}$ 服从参数 $\alpha=\sigma/\lambda$ 的瑞利分布。

4-16 设 X 是 $[-1,2]$ 上的均匀分布。求下列变换的概率密度函数。(a)$Y=\exp(-X)$；(b)$Z=X^2$。

4-17 设 X 为$[-a, a]$上的连续均匀分布随机变量。求 $Y=g(X)$的概率密度函数，其中：

$$g(x)=\begin{cases}0, & x<0\\ 3x, & 0\leqslant x\leqslant b\\ 3b, & x>b\end{cases} \quad (4\text{-}261)$$

4-18 (a)X 服从参数$\{\mu=2,\ \sigma=1\}$的高斯分布，求 $Y=\mathrm{sgn}(X)$的概率密度函数；(b)X 服从标准高斯分布和 $Z=X^3$ 时，计算 Z 的概率密度函数。

4-19 设 X 和 Y 是独立的在$[0, 1]$均匀分布的随机变量。证明对于 $V=X/Y$：

$$f_V(v)=(1/2)I_{[0,1)}(v)+(1/2v^2)I_{[1,\infty)}(v) \quad (4\text{-}262)$$

4-20 对于独立的离散随机变量 X 和 Y，推导式(4-168)中 $Z=XY$ 的概率质量函数表达式，并给出 $Z=0$ 时的结果。

4-21 设 X 是一个参数 $\{N=4,p\}$ 的二项分布随机变量。当 Y 在$[0, 2]$连续均匀分布并且与 X 独立时，计算 $Z=X+Y$ 的概率密度函数，并画出 $p=1/2$ 时的 $f_Z(z)$。

4-22 证明当 X 服从参数为 λ 的指数分布时，$Y=\exp(X)$服从参数$\{\alpha=\lambda,\ r=1\}$的帕雷托分布。

4-23 设 X_1 和 X_2 是服从几何分布的随机变量，它们相互独立并且属于$\mathcal{N}$。证明

$$P(X_1=x|X_1+X_2=y)=[1/(y-1)]I_{\{1,\cdots,y-1\}}[x] \quad (4\text{-}263)$$

是一个离散均匀分布。

4-24 当 X_1 和 X_2 独立同分布，且服从参数为 N 的卡方分布时，证明 $Y=X_1/(X_1+X_2)$是参数$\{\alpha=N/2,\ \beta=N/2\}$的贝塔随机变量。

4-25 证明 M 个独立同分布的参数为 α 的泊松随机变量，其总和也服从泊松分布但参数为 $M\alpha$。

4-26 利用逆高斯随机变量的变换导出(3-77)中的概率密度函数。

4-27 推导表 4-4 中给出的指数随机变量的变换是 Logistic 分布。(这是本章开头的图 4-1 中给出的例子)。

4-28 (a)证明当 X_1 和 X_2 是独立同分布的参数为 λ 的指数随机变量时，$Y=X_1/X_2$ 的概率密度函数为：

$$f_Y(y)=\frac{1}{(1+y)^2}I_{[0,\infty)}(y) \quad (4\text{-}264)$$

(b)利用该结论证明，$Z=\mu-\alpha\ln(X_1/X_2)$服从习题 4-27 中的 Logistic 分布。

4-29 证明当 X 服从参数 $\{\lambda=\alpha/r,r\}$ 的伽马分布时，可通过变换 $Y=\sqrt{X}$，得到概率密度函数为式(3-148)的 Nakagami 随机变量。

4-30　证明以下二项式系数总和：

$$\binom{2N}{m}=\sum_{n=0}^{m}\binom{N}{n}\binom{N}{m-n}\quad(4\text{-}265)$$

4-31　证明

$$\delta(z)=\int_0^{\infty}(\lambda/y)\delta(y)\exp(-\lambda z/y)\mathrm{d}y\quad(4\text{-}266)$$

4-32　设 X 是$\{-2,\ -1,\ 0,\ 1,\ 2\}$的等概率随机变量。定义随机向量 $\boldsymbol{Y}\triangleq[Y_1,Y_2]^{\mathrm{T}}$，其中

$$Y_1\triangleq X+2,Y_2\triangleq X^2\quad(4\text{-}267)$$

求联合概率密度函数 $f_{\boldsymbol{Y}}(\boldsymbol{y})$的表达式，并确定 Y_1 和 Y_2 是否独立。

4-33　当 X 是标准高斯随机变量时，重做习题 4-32。

4-34　设 X_1 和 X_2 是独立同分布的参数为 λ 的指数随机变量。计算 $\boldsymbol{Y}=[Y_1,Y_2]^{\mathrm{T}}$ 的联合概率密度函数，其中

$$Y_1=X_1-X_2,Y_2=X_1+X_2\quad(4\text{-}268)$$

4-35　设$\{X_1,\cdots,X_N\}$是独立同分布的参数为$\{\mu,\sigma\}$的高斯随机变量，证明式(3-211)中的$(N-1)S^2/\sigma^2$ 是自由度为 $N-1$ 的卡方分布，其中 S^2 是样本方差，与样本均值 $\overline{X}$ 有关。

顺序统计量

4-36　设$\{X_1,\cdots,X_N\}$是参数$\{\mu=0,\alpha\}$的拉普拉斯分布的样本。计算最小 $X_{(1)}$ 和最大 $X_{(N)}$ 的概率密度函数。

4-37　对于参数为 α 的泊松随机变量，重做习题 4-36。

4-38　对于参数为$\{c=0,\ \alpha\}$的柯西随机变量，重做习题 4-36。

4-39　当 X_1 和 X_2 是参数为 λ_1，λ_2 的指数随机变量时，计算最小 $X_{(1)}$ 的累积分布函数和概率密度函数。

4-40　设$\{X_1,\cdots,X_N\}$是$[\theta-1/2,\theta+1/2]$上均匀分布的样本，其中 θ 是常数。求顺序统计量$\{X_{(n)}\}$的概率密度函数的表达式。

4-41　设$\{X_1,\cdots,X_N\}$是独立同分布的$[0,\ \theta]$上均匀分布的样本。证明$\{X_{(m)},X_{(n)}\}$，$1\leqslant m\leqslant n\leqslant N$ 的联合概率密度函数为：

$$\begin{aligned}f_{X_{(m)},X_{(n)}}(x_m,x_n)=&\binom{N}{m-1,n-m-1,N-n}\\&(1/\theta^N)x_m^{m-1}(x_n-x_m)^{n-m-1}\\&(\theta-x_n)^{N-n}I_{[0,x_n]}(x_m)\\&I_{[x_m,\theta]}(x_n)\end{aligned}\quad(4\text{-}269)$$

这里使用了一个多项式的系数。

混合

4-42　设 X 是参数$\{\mu,\ \sigma\}$的高斯随机变量。证明下列参数的混合仍然是高斯的：

$$f_Y(y)=\int_{-\infty}^{\infty}a(\mu)f_X(y;\mu)\mathrm{d}\mu\quad(4\text{-}270)$$

其中 $a(\mu)$是标准高斯概率密度函数。

4-43　对于以下混合：

$$f_Y(y)=pf_{X_1}(y)+qf_{X_2}(y)\quad(4\text{-}271)$$

其中 X_1、X_2 分别是$(0,\ b]$、$(b,\ 1]$上的均匀分布，$P\in[0,1]$，$q\triangleq 1-p$。求累积分布函数 $F_Y(y)$的表达式并确定中值 m_e。

仿真作业

4-44　利用 MATLAB 中的 cdf、icdf 和式(4-87)的算法产生相关贝塔随机变量。令相关系数 $\rho=0.8$，并使用 hist3 绘制直方图。(ⅰ)二维高斯分布；(ⅱ)二维均匀分布的逆变换；(ⅲ)二维贝塔分布。设 $\alpha_1=\alpha_2=1$，$\beta_1=\beta_2=4$。使用 corrcoef 计算二维贝塔分布样本的相关系数。

4-45　重做作业 4-44 获得相关拉普拉斯随机变量。令相关系数 $\rho=-0.6$，拉普拉斯分布的参数 $\mu_1=\mu_2=0$，$\alpha_1=1$，$\alpha_2=2$。

进一步阅读

在第 2 章和第 3 章的末尾给出的参考文献里，可以发现多个随机变量和变换的额外资料。附录 A 中提供了特殊多维分布的参考资料。本章结尾的一些问题可通过下列参考文献进行进一步的研究：

混合：Johnson、Kemp 和 Kotz(2005)Johnson、Kotz 和 Balakrishnan(1995)。

顺序统计量：Lehmann(1959)，Mood、Graybill 和 Boes(1974)，Lehmann(1983)。

第5章

期望和矩

5.1 引言

随机变量的期望是其概率密度函数(pdf)的位置和中心的一个测度。从反复试验的角度来说，它是一个“预期值”，期望可以通过平均值(也称为均值)来观测。本章将讨论分布中心的其他测度，如中位数、众数以及称为矩的随机变量函数的期望。图 5-1 总结了四个能够提供随机变量位置和形状信息的重要矩。基本上，所有的期望都是一个函数对概率测度的积分。因此，它可以写成抽象概率空间 $\{\Omega,\mathcal{F},P\}$ 的勒贝格积分或随机变量 X 在 $\{\mathcal{R},\mathcal{B}(\mathcal{R}),F_X(x)\}$ 上的黎曼-斯蒂尔切斯积分。在前面的章节中，我们使用黎曼积分通过概率密度函数计算概率、导出各种随机变量函数的概率密度函数。我们通常假设概率密度函数是连续的，或者是能用狄拉克 δ 函数表示的离散、混合随机变量的概率密度函数。

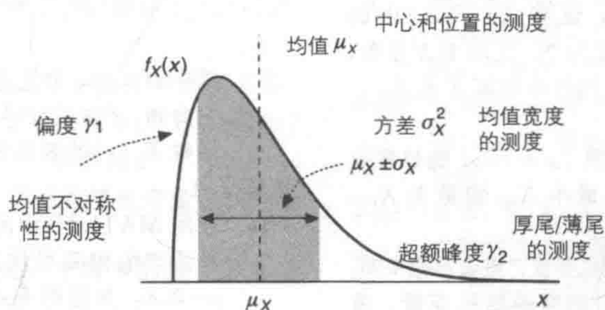


图 5-1 随机变量 X 的四个重要矩

5.2 期望与积分

首先假定抽象概率空间为 $\{\Omega,\mathcal{F},P\}$，令 X 是将 Ω 中输出映射到$\mathcal{R}$的一个随机变量。

定义(期望) 随机变量 X 的**期望**是：

$$\mathcal{E}[X]\triangleq\int_{\Omega}X(\omega)\mathrm{d}P(\omega) \tag{5-1}$$

它是函数 $X(\omega)$的勒贝格积分，P 是事件空间 $\{\Omega,\mathcal{F}\}$ 的概率测度。

附录 D 简要介绍了勒贝格积分。从 Ω 映射到$\mathcal{R}$的所有输出用上面的概率测度 P 加权后得到期望值。我们也可以按下式将$\mathcal{E}[X]$全部写成概率空间为 $\{\mathcal{R},\mathcal{B}(\mathcal{R}),F_X(x)\}$ 的随机变量的形式：

$$\mathcal{E}[X]=\int_{-\infty}^{\infty}x\mathrm{d}F_X(x) \tag{5-2}$$

其中 $F_X(x)$是 X 的累积分布函数(cdf)。该表达式是函数 $g(x)=x$ 的黎曼-斯蒂尔切斯积分，附录 D 也简要说明了这一点。在接下来的几节中，我们将探讨如何计算这些积分。首先讨论离散随机变量，然后是连续随机变量。接着，利用积分计算随机变量函数的期望，进而得到矩(如方差和偏度)和特征函数。

5.3 指示器随机变量

抽象概率空间 $\{\Omega,\mathcal{F},P\}$ 中，随机变量的一个基本类型是指示器随机变量。

定义(指示器随机变量) **指示器随机变量** $X(\omega)$是从事件 $E\in\Omega$ 到实线$\mathcal{R}$的一个映射：

$$X(w) = I_E(\omega) = \begin{cases}1, & \omega\in E\\ 0, & \text{其他}\end{cases} \tag{5-3}$$

图 5-2 给出了指示器随机变量的映射关系。观察发现 E^C 必然映射到零。除了变量 ω 是随机的，指示器随机变量与指示函数本质上是相同的，ω 由概率空间指定。它的期望是事件 E 的概率：

$$\mathcal{E}[X] = \int_\Omega I_E(\omega)\mathrm{d}P(\omega) = P(E) \tag{5-4}$$

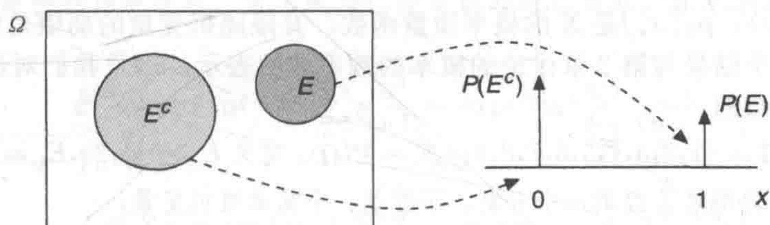


图 5-2 事件 E 的指示函数到实线$\mathcal{R}$的指示器随机变量 X 的映射

以及

$$\int_\Omega I_{E^c}(\omega)\mathrm{d}P(\omega) = P(E^c) = 1 - P(E) \tag{5-5}$$

图 5-2 中使用狄拉克 δ 函数表示这些概率。显然因为指示器随机变量只有两个输出(E 和 E^C)，X 的概率质量函数(pmf)是“成功”概率 $p=P(E)$的伯努利分布：

$$P(X=1) = P(E) = p,\quad P(X=0) = P(E^C) = 1-p \triangleq q \tag{5-6}$$

这里我们用简化的符号 X 表示 $X(\omega)$；注意 $X=x$ 实际上意味着集合$\{\omega: X(\omega)=x\}$。当掷一个骰子时，事件 $E=\{$偶数$\}$是一个 $P(E)=1/2$ 的指示器随机变量。

例 5-1 由狄利克雷函数给定的一个更为复杂的例子中，$\Omega=\mathcal{R}$，$E=\mathcal{Q}$ 是所有有理数的集合。对于指示器随机变量，有理数映射到 1，无理数(由 E^C 给定)映射到 0。事件 E 的概率是：

$$P(E) = \int_\Omega I_Q(\omega)\mathrm{d}P(\omega) = 0 \tag{5-7}$$

因此 $P(E^C)=1$：当随机选择一个实数时，我们“几乎肯定”会选择一个无理数。虽然 E 和 E^C 有无穷多的元素，无理数的基数与$\mathcal{R}$的相同(用$\beth_1$ 表示)，而$\mathcal{Q}$的基数要小得多，与整数$\mathcal{Z}$的基相同($\beth_0$ 表示)。因为没有任何有效的$\mathcal{R}$分割，黎曼积分不能用来计算概率。另一方面，因为集合使用了概率加权，可以直接利用勒贝格积分计算 $P(E)$(不同类型积分的更多讨论见附录 D)。 ◀

需要注意的是式(5-3)碰巧与第 2 章介绍的狄拉克测度相同：

$$I_E(\omega) = \delta_\omega(E) \tag{5-8}$$

但是这里我们将 $I_E(\omega)$作为概率测度 $P(\omega)$的特殊类型的随机变量。后面我们在期望的讨论中将更多地关注狄拉克测度作为概率测度。

5.4 简单随机变量

简单随机变量是指示器随机变量的概括。这里的事件不是两个，而是形成 Ω 一个分割$\mathcal{F}$中的 N 个互斥事件映射到$\mathcal{R}$的 N 个值。

定义(简单随机变量) 简单随机变量 X 具有以下形式：

$$X(\omega) = \sum_{n=1}^{N} x_n I_{E_n}(\omega) \tag{5-9}$$

其中 x_n 是$\mathcal{R}$中分配给事件 E_n 的值，$\{E_n\}$形成 Ω 的一个分割。

除了变量 ω 是随机的且由概率空间 $\{\Omega,\mathcal{F},P\}$ 指定，一个简单随机变量本质上与一个简单函数(见附录 D)相同。它遵循以下定义：

$$P(X=x_n)=P(E_n)=\int_\Omega I_{E_n}(\omega)\mathrm{d}P(\omega) \tag{5-10}$$

显然，任何具有有限个值的离散随机变量是一个简单的随机变量，因为它可以很容易地由(5-9)式表示。期望是

$$\mathcal{E}[X]=\int_\Omega\sum_{n=1}^{N}x_nI_{E_n}(\omega)\mathrm{d}P(\omega)=\sum_{n=1}^{N}x_nP(X=x_n)=\sum_{n=1}^{N}x_np_X[x_n] \tag{5-11}$$

这里利用式(5-10)，$p_X[x_n]$是 X 的概率质量函数。有限随机变量的期望是所有可能结果的概率加权和。这个结果与第 2 章讨论的概率的频率域的表示，以及我们对试验“平均结果”的直观认识是一致的。

例 5-2 令 $\Omega=\{1,2,3,4,5,6,7,8,9\}$，$\mathcal{F}=\mathcal{P}(\Omega)$。定义 $E_1=\{1,2\}$，$E_2=\{3,4,5,6\}$，$E_3=\{7,8,9\}$，它们共同形成了 Ω 的一个分割。下面是一个简单随机变量：

$$X(\omega)=\begin{cases}-1, & \omega\in E_1\\ 0, & \omega\in E_2\\ 1, & \omega\in E_3\end{cases} \tag{5-12}$$

因此 $x_1=-1$，$x_2=0$，$x_3=1$。如果 Ω 中的输出是等概率的，则 $P(X=x_1)=2/9$，$P(X=x_2)=4/9$，和 $P(X=x_3)=3/9$，进而$\mathcal{E}[X]=-1\times2/9+0\times4/9+1\times3/9=1/9$。◀

5.5 离散样本空间的期望

对于有 N 个结果的离散随机变量，期望$\mathcal{E}[X]$由式(5-11)给出。具有可数无穷个结果的离散随机变量的期望，可以通过下面介绍的简单随机变量的序列得到。首先假设 X 非负。可以直接看出简单随机变量序列存在且取极限时趋向 $X(\omega)$。例如，假设 $\Omega=\mathcal{Z}^+$ 并且令 $X(\omega)=\omega$ 为恒等映射，则 $X(\omega)$的结果为非负整数。定义以下简单随机变量序列：

$$X_n(\omega)\triangleq\begin{cases}\omega, & X(\omega)\leqslant n\\ n, & 其他\end{cases} \tag{5-13}$$

其中 $n\in\mathcal{Z}^+$。这种简单随机变量序列的结果是 0，$\{0,1\}$，$\{0,1,2\}$ 等。注意到对每个$\omega\in\Omega$，该序列是非递减的：$X_0(\omega)\leqslant X_1(\omega)\leqslant\cdots\leqslant X_n(\omega)\leqslant\cdots$。例如，当 $\omega=2$ 时，结果是 $X_0(\omega)=0$，$X_1(\omega)=0$；当$n\geqslant2$ 时，$X_n(\omega)=2$。从式(5-11)有限和的期望得出：

$$\mathcal{E}[X_n]=\sum_{m=0}^{n}mP(X=m)+nP(X>n) \tag{5-14}$$

对于每个 $\omega\in\Omega$，$\lim\limits_{n\to\infty}X_n(\omega)=X(\omega)$，期望是：

$$\begin{aligned}\mathcal{E}[X]&=\mathcal{E}[\lim_{n\to\infty}X_n]=\lim_{n\to\infty}\mathcal{E}[X_n]\\&=\lim_{n\to\infty}\sum_{m=0}^{n}mP(X=m)+\lim_{n\to\infty}nP(X>n)=\sum_{m=0}^{\infty}x_mp_X[x_m]\end{aligned} \tag{5-15}$$

式(5-15)的$\{x_m\}$表示非负整数结果。此外，为了保证期望存在，假设$\lim\limits_{n\to\infty}P(X>n)=0$ 的速度比 $n\to\infty$快。这意味着概率质量函数没有“重尾”。如果这两项以同样的速度趋于 0 和∞，则结果 $0\cdot\infty$没有定义(见附录 E)。或者，如果 $n\to\infty$的速度比 $P(X>n)\to0$ 还快，则期望无穷大。(柯西随机变量是重尾连续随机变量的一个例子，其均值没有定义且二阶矩是无穷大。它是本书讨论的具有这种情况的唯一参数随机变量。)式(5-15)的结果直接将有限随机变量的期望扩展到可数无限的随机变量。当随机变量包含负值时，可分别计算负值和正值的期望，然后它们都作为非负随机变量，最后整体期望是两个部分结果相减：$\mathcal{E}[X]=\mathcal{E}_+[X]-\mathcal{E}_-[X]$。然而对于$\mathcal{Z}$中的离散随机变量这是没必要的，可以写作：

$$\mathcal{E}[X]=\sum_{n=-\infty}^{\infty}x_n p_X[x_n] \tag{5-16}$$

例 5-3 泊松随机变量的期望是：

$$\mathcal{E}[X]=\sum_{n=0}^{\infty}n\exp(-\alpha)\frac{\alpha^n}{n!} \tag{5-17}$$

当上式为无限项之和时，可以很方便地处理该表达式：分解该式，只留下对概率质量函数的求和，最终等于1。附录E中的表E-3给出了离散概率质量函数的函数形式。另外，也可以重新排列该表达式，使它成为已知函数的级数展开。注意到求和式中第一项为0，因此：

$$\mathcal{E}[X]=\exp(-\alpha)\sum_{n=1}^{\infty}n\frac{\alpha^n}{n!}=\exp(-\alpha)\sum_{n=1}^{\infty}\frac{\alpha^n}{(n-1)!} \tag{5-18}$$

令变量 $m=n-1$，得到：

$$\mathcal{E}[X]=\exp(-\alpha)\sum_{m=0}^{\infty}\frac{\alpha^{m+1}}{m!}=\exp(-\alpha)\alpha\sum_{m=0}^{\infty}\frac{\alpha^m}{m!} \tag{5-19}$$

由于最后的求和是 $\exp(\alpha)$ 的级数展开(见附录E)，消去此项得到 $\mathcal{E}[X]=\alpha$。◀

例 5-4 几何随机变量的期望是：

$$\mathcal{E}[X]=\sum_{n=0}^{\infty}npq^n \tag{5-20}$$

该式可以借助把 n 看做乘数和幂计算。利用微分得到期望形式如下：

$$\mathcal{E}[X]=pq\frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{n=0}^{\infty}q^n\right)=pq\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{1}{1-q}\right)=pq\frac{1}{(1-q)^2} \tag{5-21}$$

由于 $q\triangleq 1-p$，最终得出 $\mathcal{E}[X]=q/p$。◀

回到期望值，它可以写作 $\mathcal{R}$ 上一个积分。离散随机变量 X 的累积分布函数可以写成单位阶跃函数的加权和：

$$F_X(x)=\sum_{n=-\infty}^{\infty}p_X[x_n]u(x-x_n) \tag{5-22}$$

其中 $\{x_n\}$ 是 X 的结果，$p_X[x_n]$ 是概率质量函数。将下式

$$\mathrm{d}F_X(x)=\sum_{n=-\infty}^{\infty}p_X[x_n]\mathrm{d}u(x-x_n) \tag{5-23}$$

代入式(5-2)得出

$$\mathcal{E}[X]=\sum_{n=-\infty}^{\infty}p_X[x_n]\int_{-\infty}^{\infty}x\mathrm{d}u(x-x_n) \tag{5-24}$$

虽然单位阶跃函数 $u(x)$ 在 $x=0$ 处没有导数，我们可以将积分写成以下狄拉克 δ 函数的形式：

$$\int_{-\infty}^{\infty}x\mathrm{d}u(x-x_n)=\int_{-\infty}^{\infty}x\delta(x-x_n)\mathrm{d}x \tag{5-25}$$

在式(5-25)里，虽然狄拉克 δ 函数是一个广义函数，但仍被视为普通函数。还要注意的是，我们可将勒贝格积分写成下式：

$$\int_{-\infty}^{\infty}x\mathrm{d}u(x-x_n)=\int_{-\infty}^{\infty}x\mathrm{d}\delta_{x_n}(x) \tag{5-26}$$

其中 $\delta_{x_n}(x)$ 是前面提到的狄拉克测度，仅当 $x=x_n$ 时为1。

由于式(5-25)中的 $\delta(x-x_n)$ 在积分里，意义是非常明确的：

$$\int_{-\infty}^{\infty}x\delta(x-x_n)\mathrm{d}x=\int_{-\infty}^{\infty}x\mathrm{d}\delta_{x_n}(x)=x_n \tag{5-27}$$

因此式(5-24)得出了与式(5-16)相同的结果。式(5-24)和式(5-25)意味着离散随机变量的

概率密度函数可以写为：

$$f_X(x)=\sum_{n=-\infty}^{\infty}p_X[x_n]\delta(x-x_n) \tag{5-28}$$

这是第 3 章中介绍的形式。虽然离散随机变量的 $f_X(x)$ 不是连续函数，但如果狄拉克 δ 函数可以理解为一个测度，那么我们就能理解如何计算包括这种 $f_X(x)$ 的积分。另外，对于函数 $g(x)$，有

$$\int_{-\infty}^{\infty}g(x)\delta(x-x_n)\mathrm{d}x=\int_{-\infty}^{\infty}g(x)\mathrm{d}\delta_{x_n}(x)=g(x_n) \tag{5-29}$$

这是后面计算离散随机变量函数的期望时要用到的知识点。

例 5-5 再次考虑概率密度函数写成以下狄拉克 δ 函数形式的几何随机变量：

$$f_X(x)=\sum_{n=0}^{\infty}pq^n\delta(x-n) \tag{5-30}$$

其期望为：

$$\mathcal{E}[X]=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x=\sum_{n=0}^{\infty}pq^n\int_{-\infty}^{\infty}x\delta(x-n)\mathrm{d}x=\sum_{n=0}^{\infty}npq^n \tag{5-31}$$

这是式(5-20)中的表达式，可以使用式(5-21)中的求微分方法计算。 ◀

5.6 连续样本空间的期望

利用简单随机变量序列也可以推导出连续随机变量的期望。这可以通过将连续随机变量近似于一个阶梯映射级数来实现(见附录 D 中勒贝格积分和连续函数的做法)。假设式(5-1)中的 $X(\omega)$ 非负并且定义下列简单随机变量：

$$X_n(\omega)\triangleq\begin{cases}m/2^n, & m/2^n\leqslant X(\omega)<(m+1)/2^n\\ 2^n, & X(\omega)\geqslant 2^n\end{cases} \tag{5-32}$$

其中 $0\leqslant m\leqslant 2^{2n}-1$。这是一个可数的随机变量，近似于原始的不可数随机变量 X。注意到对每个 $\omega\in\Omega$，$\{X_n(\omega)\}$ 是非递减的，$X_0(\omega)\leqslant X_1(\omega)\leqslant\cdots\leqslant X_n(\omega)\leqslant\cdots$。图 5-3 给出了第一个简单随机变量 $X_1(\omega)$。上述表示方法类似信号处理中用到的将连续波形转化为离散信号的量化(见第 10 章)。它也可以用于表示勒贝格积分，附录 D 总结了这一点。最后，注意到式(5-13)用类似的方法，通过恒等映射 $X(\omega)=\omega$ 和 $\Omega=\mathcal{Z}^+$ 得到可数无穷个离散随机变量。式(5-32)中的简单随机变量采用了任意非负映射 $X(\omega)$。利用非整数步进 $\{m/2n\}$ 保证当 n 增大时其趋于 0，使极限为连续随机变量，如定理 5-1 中所描述的。

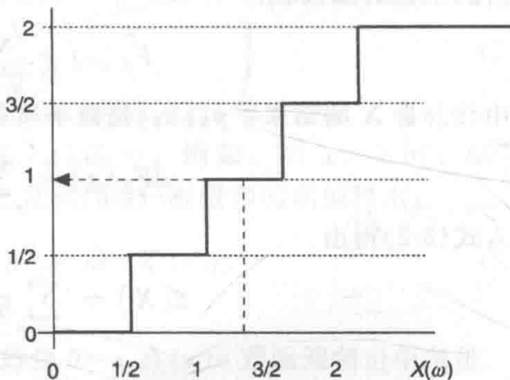


图 5-3 由简单随机变量 $X_1(\omega)$ 近似 $X(\omega)$

定理 5-1 一个简单的随机变量序列的极限是连续非负随机变量。

证明：利用式(5-9)中使用的符号表示这个简单的随机变量，$X(\omega)$ 的第 n 个近似改写如下：

$$X_n(\omega)=\sum_{m=0}^{2^{2n}-1}(m/2^n)I_{\{m/2n\leqslant X(\omega)<(m+1)/2n\}}(\omega)+2^nI_{\{X(\omega)\geqslant 2^n\}}(\omega) \tag{5-33}$$

观察发现

$$\sup_{m\geqslant n}|X(\omega)-X_m(\omega)|\leqslant 2^{-n} \tag{5-34}$$

其中 2^{-n} 是每个区间的宽度(不包括趋于∞的阶梯函数的最后一个台阶)。对于每个 $\omega\in\Omega$，

阶梯近似是非递减的。随着n的增加，步进减少，台阶范围扩展到覆盖$\mathcal{R}^+$。当 $n\to\infty$时，等式右边趋于 0，近似值收敛于原始随机变量：$X_n(\omega)\to X(\omega)$。(由 n 索引的随机变量序列的收敛性将在第 7 章进一步讨论)。

通过这个近似，连续随机变量(非负)的期望是：

$$\mathcal{E}[X]=\lim_{n\to\infty}\sum_{m=0}^{2^{2n}-1}(m/2^n)P(m/2^n\leqslant X(\omega)<(m+1)/2^n)+\lim_{n\to\infty}2^nP(X(\omega)\geqslant 2^n) \quad (5\text{-}35)$$

$$=\lim_{n\to\infty}\sum_{m=0}^{2^{2n}-1}x_{m,n}P(X\in[m/2^n,(m+1)/2^n]) \quad (5\text{-}36)$$

其中 $x_{m,n}\triangleq m/2^n$，并且为了去掉最后一项，假定 $2^{-n}\to\infty$的速度比 $P(X(\omega)\geqslant 2^n)\to 0$ 的速度慢。再次注意 $x_{m,n}$不像式(5-16)中的离散随机变量 x_n，它不是整数值。期望是概率测度为 P_X 的非负 x 的勒贝格积分：

$$\mathcal{E}[X]=\int_{-\infty}^{\infty}x\mathrm{d}P_X(x) \quad (5\text{-}37)$$

任意连续随机变量的期望通过正结果期望减去负结果期望获得：$\mathcal{E}[X]=\mathcal{E}_+[X]-\mathcal{E}_-[X]$。

例 5-6 设 X 是指数随机变量，累积分布函数如下：

$$F_X(x)=[1-\exp(-\lambda x)]I_{[0,\infty)}(x) \quad (5\text{-}38)$$

简单随机变量 $X_1(\omega)$的期望为：

$$\mathcal{E}[X_1]=0P(0\leqslant X<1/2)+(1/2)P(1/2\leqslant X<1)$$
$$1P(1\leqslant X<3/2)+(3/2)P(3/2\leqslant X<2)+2P(X\geqslant 2) \quad (5\text{-}39)$$

利用累积分布函数，可得：

$$F_X(a)-F_X(b)=\exp(-\lambda b)-\exp(-\lambda a) \quad (5\text{-}40)$$

因此

$$\mathcal{E}[X_1]=(1/2)[\exp(-\lambda/2)-\exp(-\lambda)]+[\exp(-\lambda)-\exp(-3\lambda/2)]$$
$$+(3/2)[\exp(-3\lambda/2)-\exp(-2\lambda)]+2\exp(-2\lambda) \quad (5\text{-}41)$$

当 $\lambda=1$ 时，利用式(5-41)得出$\mathcal{E}[X_1]=0.6664$。随着 n 的增加重新计算期望，很容易得出$\mathcal{E}[X_n]\to\mathcal{E}[X]=1$。特别注意到

$$P(X\geqslant 2^n)=1-F_X(2^n)=\exp(-2^n\lambda) \quad (5\text{-}42)$$

从而：

$$\lim_{n\to\infty}2^n\exp(-2^n\lambda)=0 \quad (5\text{-}43)$$

即式(5-35)最后一项的概率趋于零的速度要快于 $2^{-n}\to\infty$。◀

随机变量 $X(\omega)$从 $\{\Omega,\mathcal{F},\mathcal{P}\}$ 映射到 $\{\mathcal{R},\mathcal{B}(\mathcal{R}),F_X(x)\}$，期望可以通过在$\mathcal{R}$上进行黎曼-斯蒂尔切斯积分得到：

$$\mathcal{E}[X]=\int_{-\infty}^{\infty}x\mathrm{d}F_X(x) \quad (5\text{-}44)$$

其中 $F_X(x)$是 X 的累积分布函数。如果 $F_X(x)$处处可导，那么 $\mathrm{d}F_X(x)=f_X(x)\mathrm{d}x$，其中 $f_X(x)$是概率密度函数，则：

$$\mathcal{E}[X]=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x \quad (5\text{-}45)$$

这个表达式定义的期望更加常见，是式(5-15)离散随机变量期望在连续随机变量的对应的表达式；这里积分“替换”求和。对于$\mathcal{R}$或抽象的样本空间 Ω 中的混合随机变量，分别利用黎曼-斯蒂尔切斯和勒贝格积分计算期望。附录 D 中提供了这方面的例子。

例 5-7 因为高斯随机变量关于参数 μ 对称，因此其期望很容易计算。但是要注意，这种对称性并不总是保证得到明确的期望，稍后将借助柯西随机变量说明这一点。当函数关于某个 μ 对称时，我们可以得出 $f_X(x+\mu)=f_X(-x+\mu)$。将 $f_X(x)$乘以一个 μ 的奇函

数，则积分为零。令奇函数为 $g(x)=\mu-x$。则：

$$\int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x = \int_{-\infty}^{\infty} (x-\mu) f_X(x) \mathrm{d}x = 0 \tag{5-46}$$

假定 $g(x)f_X(x)$ 的积分在每个区间是有限值，重新组合上面表达式，得到最终结果：

$$\mathcal{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x = \mu \int_{-\infty}^{\infty} f_X(x) \mathrm{d}x = \mu \tag{5-47}$$ ◀

例 5-8 通常利用分段积分(见附录 D)来计算指数随机变量的期望(以及后面的其他矩)：

$$\int_a^b u \mathrm{d}v = uv \Big|_a^b - \int_a^b v \mathrm{d}u \tag{5-48}$$

设 $u \equiv \lambda x$，$\mathrm{d}v \equiv \exp(-\lambda x)$，因此

$$\begin{aligned}\mathcal{E}[X] &= \int_0^{\infty} x\lambda \exp(-\lambda x) \mathrm{d}x \\ &= \lambda x(-1/\lambda)\exp(-\lambda x)\Big|_0^{\infty} - \int_0^{\infty} \lambda(-1/\lambda)\exp(-\lambda x)\mathrm{d}x \\ &= 0 + \int_0^{\infty} \exp(-\lambda x)\mathrm{d}x = 1/\lambda \end{aligned} \tag{5-49}$$ ◀

例 5-9 位置参数 $c=0$ 的柯西随机变量，其期望是：

$$\begin{aligned}\mathcal{E}[X] &= (\alpha/\pi)\int_{-\infty}^{\infty} \frac{x}{x^2+\alpha^2}\mathrm{d}x \\ &= (\alpha/2\pi)\ln(x^2+\alpha^2)\Big|_{-\infty}^{\infty} = \infty - \infty \end{aligned} \tag{5-50}$$

虽然 x 是奇函数，$f_X(x)$ 是偶数，但由于积分结果在每个端点是无限的：$\infty-\infty$ 没有定义(见附录 B)，因此 $\mathcal{E}[X]$ 不存在。 ◀

5.7 期望的总结

本节总结了随机变量期望的主要结果。抽象概率空间 $\{\Omega,\mathcal{R},P\}$ 中随机变量 X 的期望是：

$$\mathcal{E}[X] = \int_{\Omega} X(\omega)\mathrm{d}P(\omega) \quad (勒贝格积分) \tag{5-51}$$

其中 $P(\omega)$ 是概率测度。概率空间 $\{\mathcal{R},\mathcal{B}(\mathcal{R}),F_X(x)\}$ 中随机变量 X 的期望是以下积分：

$$\mathcal{E}[X] = \int_{-\infty}^{\infty} x\mathrm{d}F_X(x) \quad (黎曼\text{-}斯蒂尔切斯积分) \tag{5-52}$$

其中 x 是被积函数，累积分布函数 $F_X(x)$ 是积分变量。如果 $F_X(x)$ 处处可微，那么可以将 $\mathrm{d}F_X(x)=f_X(x)\mathrm{d}x$ 代入最后一个表达式，表示成普通的黎曼积分：

$$\mathcal{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x \quad (黎曼积分) \tag{5-53}$$

其中 $f_X(x)$ 是概率密度函数(且 $xf_X(x)$ 是被积函数)。如果狄拉克 δ 函数可以表示概率密度函数，该结果可以用于离散随机变量(以及混合随机变量)。概率质量函数为 $p_X[x]$ 的离散随机变量，期望的计算方法如下：

$$\mathcal{E}[X] = \int_{-\infty}^{\infty} x\sum_{n=-\infty}^{\infty} p_X[x_n]\delta(x-x_n)\mathrm{d}x = \sum_{n=-\infty}^{\infty} x_n p_X[x_n] \quad (无限和) \tag{5-54}$$

这里利用了狄拉克 δ 函数的筛选特性。本书的后续部分，会用到式(5-53)和式(5-54)中的表达式。可以将这些结果推广到随机变量 X 函数的期望以及随机向量函数的期望(本章稍后将讨论)。

最后，注意到当计算多个随机变量的期望时，有时搞不清究竟使用哪个随机变量的分布。出于这一点，我们可以更确切地将 X 的期望记为 $\mathcal{E}_X[x]$，其中下标表示求和或积分是针对 X 的概率质量函数或概率密度函数。下标符号在后面的条件期望中将特别有用。当使用的分布显而易见时，可以利用速记符号 $\mathcal{E}[X]$ 简化表达式。

5.8 均值的函数观点

期望$\mathcal{E}[X]$也称随机变量X的均值：它是概率密度函数中心和位置的一个测度。通过将均值看成两个函数乘积的面积，有利于更加深入地研究它。观察到式(5-53)中的被积函数是概率密度函数$f_X(x)$(当然总是非负的)与x(关于原点是奇函数)的乘积。令$f_X(x)$是参数$\{\mu=0, \sigma=1\}$的标准高斯概率密度函数。图5-4a分别画出了x，$f_X(x)$以及$x\in[-5\sigma,5\sigma]$时二者的乘积。由于$g(x)=x$是奇函数，当$x<0$时，它翻转了乘积$xf_X(x)$中的$f_X(x)$。当对结果进行积分时，$xf_X(x)$的正负区域相互抵消。当$\mu=0$时，由于$f_X(x)$是关于原点的偶函数，正负区域完全抵消。我们也可以将乘积的累积区域记作：

$$A(x)\triangleq\int_{-\infty}^{x} v f_X(v)\mathrm{d}v \tag{5-55}$$

x在$f_X(x)$的定义域中可变。(通过将x作为$f_X(x)$的权值，$A(x)$可以看作加权后的"累积分布函数"。)显然，$\mathcal{E}[X]=\lim\limits_{x\to\infty}A(x)$。

图5-4b给出了$\mu=1$，将高斯概率密度函数右移的结果。在这种情况下，$xf_X(x)$的正负区域没有完全抵消，这是由于概率密度函数进行了右移，整体面积是正的。于是我们看到，通过将$f_X(x)$乘以通过x轴原点的函数，乘积$xf_X(x)$可以反映$f_X(x)$相对于原点的位置。图5-5给出了指数随机变量的类似结果，在$x<0$时随机变量为0，因此累积面积$A(x)$不会是负值。

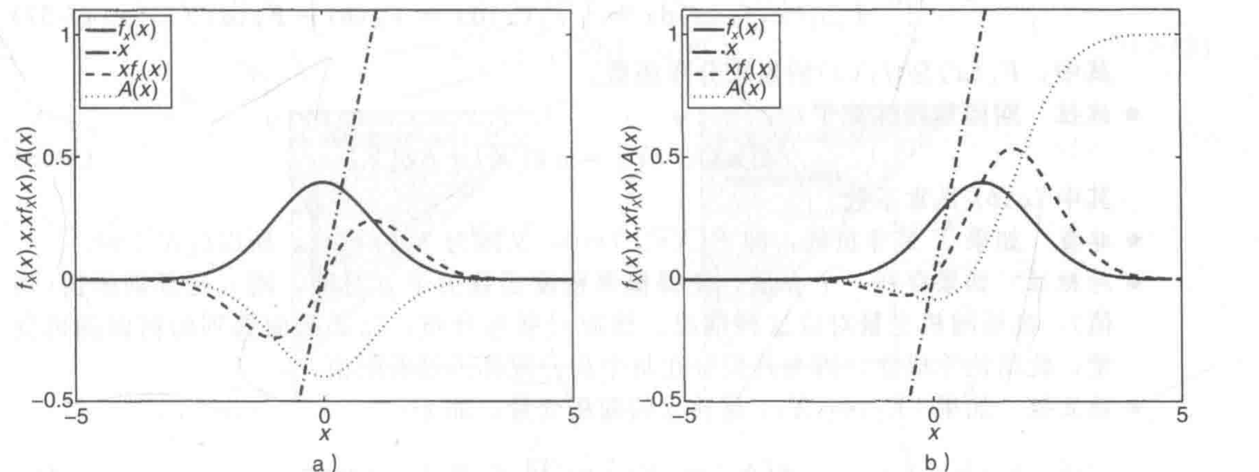


图5-4 均值的函数观点，给出了高斯概率密度函数的$f_X(x)$，x，$xf_X(x)$和$A(x)$。a) $\mu=0$、$\sigma=1$；b) $\mu=1$、$\sigma=1$

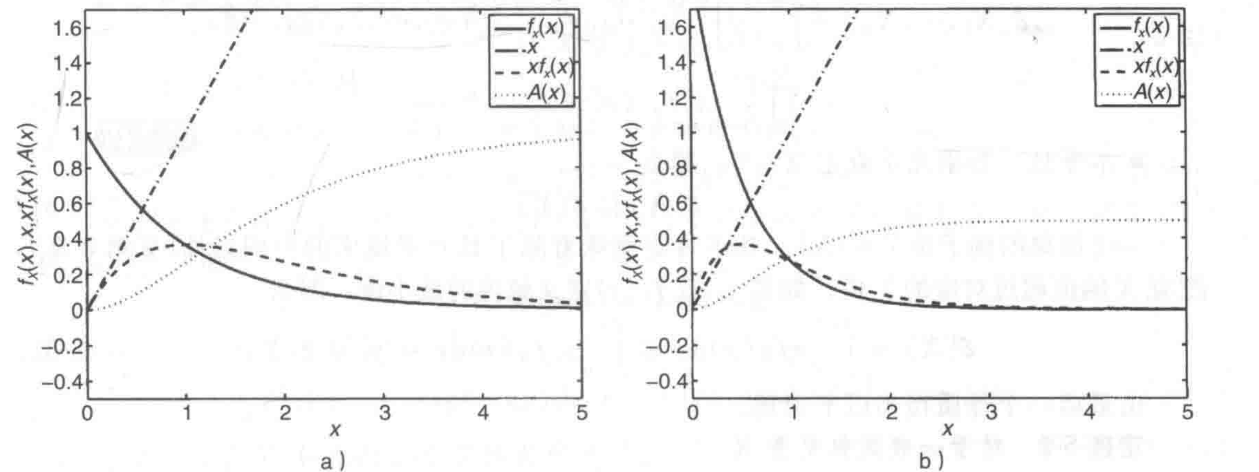


图5-5 均值的函数观点，给出了指数概率密度函数的$f_X(x)$，x，$xf_X(x)$和$A(x)$。a) $\lambda=2$；b) $\lambda=1$

从这个函数观点观察到的一个重要特征是，当 $f_X(x)$乘以 $g(x)=x$ 并且积分后，更多地关注大的 x 值。此性质解释了为什么柯西概率密度函数具有“重尾”但是没有均值(没有定义它)。将乘积 $xf_X(x)$简单地看作一个函数并积分，概率密度函数中心两侧的区域可能不是有限的，而这正好对应柯西概率密度函数的情况。$f_X(x)$的重尾容许大的 x 值(通过函数 $g(x)=x$)主导：$f_X(x)$的衰减速度没有$|x|$增大的速度快。用类推的方法，这种特点类似于不收敛的无穷级数。因此，即使柯西概率密度函数上确实有一个中心点(由中位数和众数给定)，由于前面提到了$\infty-\infty$是没有定义的，因此均值不存在。这个结果表明，奇函数和偶函数乘积的积分不一定是 0。

从这个函数观点我们得出结论，奇函数提供了分布中心的一个测度。这方面的例子包括所有幂次是奇数的高阶矩，如$\mathcal{E}[X^3]$，$\mathcal{E}[X^5]$等。后面我们使用一个类似的函数描述来解释为什么方差(二阶中心矩)是分布宽度的测度。

5.9　期望的性质

期望的下列性质适用于离散、连续和混合随机变量。

- **指示函数**　指示函数的期望是一个概率：

$$\mathcal{E}[I_{[a,b]}(x)]=p(a\leqslant X\leqslant b) \tag{5-56}$$

很容易得出：

$$\int_{-\infty}^{\infty} I_{[a,b]}(x)f_X(x)\mathrm{d}x=\int_a^b f_X(x)\mathrm{d}x=F_X(b)-F_X(a) \tag{5-57}$$

其中，$F_X(x)$是 $f_X(x)$的累积分布函数。

- **线性**　期望是线性算子：

$$\mathcal{E}[aX+bY]=a\,\mathcal{E}[X]+b\,\mathcal{E}[Y] \tag{5-58}$$

其中$\{a,b\}$是常系数。

- **非负**　如果 X 是非负的，即 $P(X<0)=0$，又因为 $f_X(x)\geqslant 0$，所以$\mathcal{E}[X]\geqslant 0$。
- **对称性**　如果存在一个 μ 值，使得概率密度函数关于 μ 对称，则 μ 是预期的值(均值)。高斯随机变量对应这种情况。然而对某些分布，比如前面提到的柯西随机变量，此结论不成立，因为其积分在每个积分限都不是有限值。
- **独立性**　如果$\{X_1,\cdots,X_N\}$是独立的随机变量，那么：

$$\mathcal{E}[X_1,\cdots,X_N]=\prod_{n=1}^{N}\mathcal{E}[X_n] \tag{5-59}$$

这是因为联合概率可以写成边缘概率密度函数的乘积：

$$\begin{aligned}\mathcal{E}[X_1,\cdots,X_N]&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_1\cdots x_N f_{X_1,\cdots,X_N}(x_1,\cdots,x_N)\mathrm{d}x_1\cdots\mathrm{d}x_N\\&=\prod_{n=1}^{N}\int_{-\infty}^{\infty}x_n f_{X_n}(x_n)\mathrm{d}x_n\end{aligned} \tag{5-60}$$

- **不等性**　如果几乎肯定 $X\leqslant Y$，那么

$$\mathcal{E}[X]\leqslant\mathcal{E}[Y] \tag{5-61}$$

一个简单的例子是 $Y=|X|$。几乎肯定意味着除了具有零概率的个别点外(见第 2 章)，没有 X 的值超过对应的 Y 值。如果 y_o 是 $f_Y(y)$定义域内的最小值，那么

$$\mathcal{E}[X]=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x\leqslant\int_{-\infty}^{\infty}y_o f_X(x)\mathrm{d}x=y_o\leqslant\mathcal{E}[Y] \tag{5-62}$$

由最后一个性质得出以下定理。

定理 5-2　对于一般随机变量 X

$$\mathcal{E}[X]\leqslant|\mathcal{E}[X]|\leqslant\mathcal{E}[|X|] \tag{5-63}$$

证明：第一个不等式是显而易见的。对于第二个不等式

$$|\mathcal{E}[X]| \leqslant \int_{-\infty}^{\infty} |xf_X(x)|\mathrm{d}x = \int_{-\infty}^{\infty} |x| f_X(x)\mathrm{d}x = \mathcal{E}[|X|] \tag{5-64}$$

因为对所有 x，$f_X(x)$非负。该结果也可以通过式(5-61)中的不等式性质得到，其中$Y=|X|$，要注意 $X \leqslant |X|$，$-X \leqslant |X|$。

定理 5-3 对于非负连续随机变量 X

$$\mathcal{E}[X] = \int_0^{\infty} P(X \geqslant x)\mathrm{d}x \tag{5-65}$$

对于非负离散随机变量 X

$$\mathcal{E}[X] = \sum_{n=1}^{\infty} P(X \geqslant n) \tag{5-66}$$

其中求和从 $n=1$(不是 $n=0$)开始。

证明：这里仅提供连续情况的证明。代入概率密度函数，得到：

$$\int_0^{\infty} P(X \geqslant x)\mathrm{d}x = \int_0^{\infty}\left[\int_X^{\infty} f_X(v)\mathrm{d}v\right]\mathrm{d}x \tag{5-67}$$

图 5-6a 给出了 X-V 平面的积分区域(阴影部分)。先关于 x 进行积分，需要改变积分限以进行水平积分，如图 5-6b 所示。这里给出的是对一个常数进行内积分：

$$\int_0^{\infty} P(X \geqslant x)\mathrm{d}x = \int_0^{\infty}\left[\int_0^{v} f_X(v)\mathrm{d}x\right]\mathrm{d}v = \int_0^{\infty} f_X(v)\left[\int_0^{v}\mathrm{d}x\right]\mathrm{d}v$$

$$= \int_0^{\infty} vf_X(v)\mathrm{d}v = \mathcal{E}[X] \tag{5-68}$$

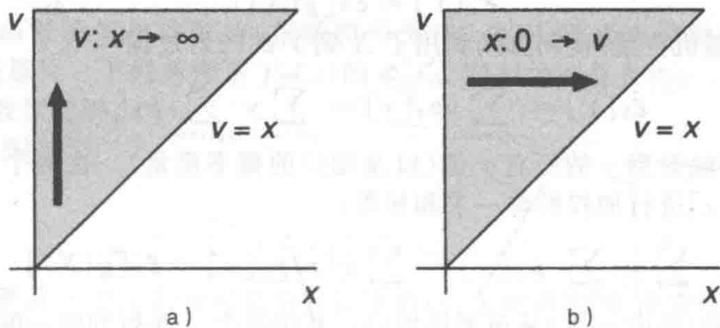


图 5-6 式(5-67)的积分区域。a) 先关于 v 进行垂直积分；b) 先关于 x 进行水平积分

离散情况的证明要求交换两个求和的顺序(见习题 5-7)。

- 马尔可夫不等式(5-67)与马尔可夫不等式有关(见附录 F)：

$$P(X \geqslant x_0) \leqslant \mathcal{E}[X]/x_0 \tag{5-69}$$

其中 $x_0 > 0$ 是常数。

例 5-10 考虑具有以下累积分布函数的指数随机变量 X

$$F_X(x) = [1 - \exp(-\lambda x)]I_{\mathcal{R}^+}(x) \tag{5-70}$$

对于连续随机变量，$P(X>x)=P(X\geqslant x)$(可以包括或不包括等号：它们都等于$1-F_X(x)$)。从而

$$P(X \geqslant x) = \exp(-\lambda x) \tag{5-71}$$

因为 X 非负，从式(5-65)可以得出：

$$\mathcal{E}[X] = \int_0^{\infty} \exp(-\lambda x)\mathrm{d}x = (-1/\lambda)\exp(-\lambda x)\Big|_0^{\infty} = 1/\lambda \tag{5-72}$$

直接计算$\mathcal{E}[X]$要像例 5-8 那样进行分部积分。用式(5-67)更加简单，因为$P(X\geqslant x)$可看做一个积分，并可通过累积分布函数得到。 ◀

例 5-11 对于具有以下累积分布函数的几何随机变量 X

$$F_X(x) = (1-q^{x+1})I_{\mathbb{Z}^+}[x] \tag{5-73}$$

期望是：

$$\begin{aligned}\mathcal{E}[X] &= \sum_{n=1}^{\infty} P(X \geqslant n) = \sum_{n=1}^{\infty}[1-P(X<n)] \\ &= \sum_{n=1}^{\infty}[1-F_X(n)+pq^n]\end{aligned} \tag{5-74}$$

代入累积分布函数并使用附录 E 中给出的无穷级数的结果，得出：

$$\mathcal{E}[X] = (p+q)\sum_{n=0}^{\infty} q^n - 1 = 1/(1-q) - 1 = q/p \tag{5-75}$$

这是例 5-4 直接计算$\mathcal{E}[X]$得出的均值。观察发现，因为 $P(X \geqslant n)$里有等号，我们需要包含 n 处的概率质量(式(5-74)的最后一行)。减去包含等号的累积分布函数，再将 pq^n 加入表达式。$n=0$ 对应的部分(等于 1)也要被减去，因为要保证求和从 0 开始以及得到无限和的闭合表达式。◀

5.10 函数的期望

随机变量 X 的函数 $Y=g(X)$的期望可以由两种方式计算：(ⅰ)推导概率密度函数 $f_Y(y)$后计算$\mathcal{E}_Y(y)$；(ⅱ)直接基于 $f_X(x)$计算$\mathcal{E}_X[g(X)]$。在下面的定理中，我们证明了这两种方法是等价的。

定理 5-4 随机变量 X 的函数 $Y=g(X)$的期望是：

$$\mathcal{E}_Y[Y] = \mathcal{E}_X[g(X)] \tag{5-76}$$

证明：(离散随机变量)证明直接利用了 X 到 Y 的映射过程

$$\mathcal{E}_Y[Y] = \sum_{y=-\infty}^{\infty} y p_Y[y] = \sum_{y=-\infty}^{\infty} y \sum_{x:g(x)=y} p_X[x] \tag{5-77}$$

内求和利用了映射到 y 的所有 x 值(以及相应的概率质量)。这两个求和与对 $g(x)$的所有可能值用 $p_X[x]$进行加权的单一求和相等：

$$\sum_{y=-\infty}^{\infty} y \sum_{x:g(x)=y} p_X[x] = \sum_{x=-\infty}^{\infty} g(x)p_X[x] = \mathcal{E}_X[g(X)] \tag{5-78}$$

(连续随机变量)假设 $g(X)$是单调递增的，其中每个 x 映射到唯一的 y。从而：

$$\mathcal{E}_Y[Y] = \int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y = \int_{-\infty}^{\infty} y f_X(g^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|\mathrm{d}y \tag{5-79}$$

代入 $y=g(x)$和 $x=g^{-1}(y)$，得出：

$$\mathcal{E}_Y[Y] = \int_{-\infty}^{\infty} g(x)f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|\mathrm{d}y = \int_{-\infty}^{\infty} g(x)f_X(x)\mathrm{d}x = \mathcal{E}_X[g(X)] \tag{5-80}$$

其中单调递增函数的 $\mathrm{d}x/\mathrm{d}y>0$。更通用函数的方法与第 4 章随机变量变换使用的方法相同。

例 5-12 设 X 是参数为 $\{\mu,\sigma\}$ 的高斯随机变量。从第 4 章可知，$Y=g(X)=\exp(X)$是相同参数的对数正态分布。关于 X 计算期望：

$$\begin{aligned}\mathcal{E}_X[g(X)] &= \int_{-\infty}^{\infty} \exp(x)\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/2\sigma^2)\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp(x-(x-\mu)^2/2\sigma^2)\mathrm{d}x\end{aligned} \tag{5-81}$$

对指数幂次的平方项进行整理，得到：

$$\mathcal{E}_X[g(X)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp(-([x-(\mu+\sigma^2)]^2-(\mu+\sigma^2)^2+\mu^2)/2\sigma^2)\mathrm{d}x \tag{5-82}$$

提出常数项，得出：

$$\mathcal{E}_X[g(X)] = \exp(\mu + \sigma^2/2) \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp(-(x-(\mu+\sigma^2))^2/2\sigma^2)\mathrm{d}x$$

$$= \exp(\mu + \sigma^2/2) \tag{5-83}$$

因为被积函数是参数为$\{\mu+\sigma^2, \sigma\}$的高斯概率密度函数，最后一个积分值等于1。这个结果可以通过直接计算对数正态概率密度函数的均值进行验证：

$$\mathcal{E}_Y[Y] = \int_0^{\infty} \frac{1}{y\sqrt{2\pi\sigma^2}} y\exp(-(\ln(y)-\mu)^2/2\sigma^2)\mathrm{d}y$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp(x)\exp(-(x-\mu)^2/2\sigma^2)\mathrm{d}x = \mathcal{E}_X[g(X)] \tag{5-84}$$

这里 $\mathrm{d}y = \exp(x)\mathrm{d}x$ 利用了变换 $x = \ln(y)$，从而积分限变为$\{-\infty, \infty\}$。 ◀

5.11 特征函数

特征函数是随机变量一个特定复函数的期望。在系统地确定随机变量矩和变量函数(如第4章中两个独立随机变量的总和)时，它是非常有用的。

定义(特征函数) 随机变量 X 的**特征函数**(Characteristic Function，CF)是：

$$\Phi_X(\omega) \triangleq \mathcal{E}[\exp(\mathrm{j}\omega X)] \tag{5-85}$$

其中 $\mathrm{j} = \sqrt{-1}, \omega \in [0, 2\pi]$。

将 $g(X) = \exp(\mathrm{j}\omega X)$代入式(5-80)，得出：

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)\exp(\mathrm{j}\omega x)\mathrm{d}x \tag{5-86}$$

它类似于概率密度函数的傅里叶变换(指数符号为正，而不是通常傅里叶变换定义里的负)。像傅里叶变换那样，下列条件是 $f_X(x)$的 $\Phi_X(\omega)$存在的充分条件：

- 不连续点的数目有限。
- 变化有限(见附录B)。
- 绝对可积：

$$\int_{-\infty}^{\infty} |f_X(x)|\mathrm{d}x < \infty \tag{5-87}$$

需要注意的是附录A总结的分布都有特征函数(甚至柯西)，因为有些分布的特征函数很复杂，这里没有给出。

令$\{X_1, \cdots, X_N\}$是独立的随机变量，其概率密度函数可以是任意的。这些随机变量总和的特征函数可以写为乘积的形式：

$$\Phi_{X_1+\cdots+X_N}(\omega) = \mathcal{E}\left[\exp\left(\mathrm{j}\omega\sum_{n=1}^{N} X_n\right)\right] = \mathcal{E}\left[\prod_{n=1}^{N}\exp(\mathrm{j}\omega X_n)\right] \tag{5-88}$$

计算期望值最初利用了$\{X_n\}$的联合概率密度。由于随机变量独立，联合概率密度函数可以写成 N 个边缘概率密度函数乘积的形式，得到：

$$\Phi_{X_1+\cdots+X_N}(\omega) = \prod_{n=1}^{N}\mathcal{E}[\exp(\mathrm{j}\omega X_n)] = \prod_{n=1}^{N}\Phi_{X_n}(\omega) \tag{5-89}$$

其中乘积中的第 n 个期望与 X_n 的概率密度函数有关。

下面的定理描述了特征函数矩的矩母性质(本章稍后将介绍高阶矩)。

定理5-5 利用特征函数生成随机变量 X 的非中心(原始)矩方法如下：

$$\mathcal{E}[X^n] = (-\mathrm{j})^n \frac{\mathrm{d}^n}{\mathrm{d}\omega^n}\Phi_X(\omega)\Big|_{\omega=0} \tag{5-90}$$

证明：(功率级数法)利用指数信号的麦克劳林级数展开指数信号(见附录E)，得到：

$$\mathcal{E}[\exp(\mathrm{j}\omega X)] = \mathcal{E}\left[1 + \mathrm{j}\omega X + \frac{(\mathrm{j}\omega X)^2}{2!} + \cdots + \frac{(\mathrm{j}\omega X)^k}{k!} + \frac{(\mathrm{j}\omega X)^{k+1}}{(k+1)!} + \cdots\right] \tag{5-91}$$

使用等式：

$$(-\mathrm{j})^n \frac{\mathrm{d}^n\Phi_X(\omega)}{\mathrm{d}\omega^n} = \frac{\mathrm{d}^n\Phi_X(\omega)}{\mathrm{d}(\mathrm{j}\omega)^n} \tag{5-92}$$

微分 n 次后展开式的前$(n-1)$项消失：

$$\frac{\mathrm{d}^n}{\mathrm{d}(\mathrm{j}\omega)^n}\Phi_X(\omega) = \mathcal{E}[X^n + \mathrm{j}\omega X^{n+1} + (1/2)(\mathrm{j}\omega)^2 X^{n+2} + \cdots] \tag{5-93}$$

将 $\omega=0$ 代入后，得到以下结果，除了与 ω 无关的第一项外，其余项消去：

$$\frac{\mathrm{d}^n}{\mathrm{d}(\mathrm{j}\omega)^n}\Phi_X(\omega)\bigg|_{\omega=0} = \mathcal{E}[X^n + 0 + \cdots] = \mathcal{E}[X^n] \tag{5-94}$$

（定义法）也可以通过特征函数的定义：

$$\begin{aligned}\frac{\mathrm{d}^n}{\mathrm{d}(\mathrm{j}\omega)^n}\mathcal{E}[\exp(\mathrm{j}\omega X)]\bigg|_{\omega=0} &= \int_{-\infty}^{\infty}\frac{\mathrm{d}^n}{\mathrm{d}(\mathrm{j}\omega)^n}\exp(\mathrm{j}\omega x)f_X(x)\mathrm{d}x\,\Big|_{\omega=0} \\ &= \int_{-\infty}^{\infty} x^n\exp(\mathrm{j}\omega x)f_X(x)\mathrm{d}x\,\Big|_{\omega=0}\end{aligned} \tag{5-95}$$

在 $\omega=0$ 处消去表达式积分项里的指数部分，得到：

$$\frac{\mathrm{d}^n}{\mathrm{d}(\mathrm{j}\omega)^n}\mathcal{E}[\exp(\mathrm{j}\omega X)]\bigg|_{\omega=0} = \int_{-\infty}^{\infty} x^n f_X(x)\mathrm{d}x = \mathcal{E}[X^n] \tag{5-96}$$

定理得证。

该定理还表明，随机变量的矩共同描述了概率密度函数：所有矩组合起来包含了与概率密度函数相同的“信息”。

例 5-13 指数随机变量的特征函数为：

$$\begin{aligned}\Phi_X(\omega) &= \int_0^{\infty}\lambda\exp(-\lambda x)\exp(\mathrm{j}\omega x)\mathrm{d}x \\ &= \lambda\int_0^{\infty}\exp(-x(\lambda-\mathrm{j}\omega))\mathrm{d}x = \frac{\lambda}{\lambda-\mathrm{j}\omega}\end{aligned} \tag{5-97}$$ ◀

一阶矩（均值）的计算方法如下：

$$\mathcal{E}[X] = \frac{\mathrm{d}}{\mathrm{d}(\mathrm{j}\omega)}\left(\frac{\lambda}{\lambda-\mathrm{j}\omega}\right)\bigg|_{\omega=0} = \frac{\lambda}{(\lambda-\mathrm{j}\omega)^2}\bigg|_{\omega=0} = 1/\lambda \tag{5-98}$$

例 5-14 二项随机变量的特征函数为：

$$\begin{aligned}\Phi_X(\mathrm{j}\omega) &= \int_{-\infty}^{\infty}\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}\exp(\mathrm{j}\omega x)\delta(x-n)\mathrm{d}x \\ &= q^N\sum_{n=0}^{N}\binom{N}{n}[(p/q)\exp(\mathrm{j}\omega)]^n\end{aligned} \tag{5-99}$$

其中 $q \triangleq 1-p$，$x=n$ 处的狄拉克 δ 函数导致 $\exp(\mathrm{j}\omega x)$ 积分后为 $\exp(\mathrm{j}\omega n)$。使用二项式公式：

$$(x+y)^N = \sum_{n=0}^{N}\binom{N}{n}x^{N-n}y^n \tag{5-100}$$

可以得到：

$$\Phi_X(\mathrm{j}\omega) = q^N[1+(p/q)\exp(\mathrm{j}\omega)]^N = [q+p\exp(\mathrm{j}\omega)]^N \tag{5-101}$$

平均值计算方法如下：

$$\mathcal{E}[X] = \frac{\mathrm{d}}{\mathrm{d}(\mathrm{j}\omega)}[q+p\exp(\mathrm{j}\omega)]^N\,\Big|_{\omega=0} = N[q+p\exp(\mathrm{j}\omega)]^{N-1}p\exp(\mathrm{j}\omega)\,\Big|_{\omega=0} = Np \tag{5-102}$$

代入 $\omega=0$ 后，中括号里的值为 $q+p=1$。 ◀

注意因为类似于离散时间序列的离散时间傅里叶变换（DTFT），离散概率质量函数的 $\Phi_X(\mathrm{j}\omega)$ 参数中包括了 j。写成概率质量函数的形式如下：

$$\Phi_X(\mathrm{j}\omega)=\sum_{n=-\infty}^{\infty}p_X[n]\exp(\mathrm{j}\omega n) \tag{5-103}$$

例 5-15 柯西概率密度函数的特征函数为：

$$\begin{aligned}\Phi_X(\omega)&=\int_{-\infty}^{\infty}\frac{\alpha/\pi}{(x-c)^2+\alpha^2}\exp(\mathrm{j}\omega x)\mathrm{d}x\\&=\exp(\mathrm{j}\omega c)\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{2\alpha}{v^2+\alpha^2}\exp(\mathrm{j}\omega v)\mathrm{d}v\end{aligned} \tag{5-104}$$

这里将第二个表达式的变量变成 $v=x-c$。从表中的傅里叶变换对（见附录 C）得到积分（包括因子 $1/2\pi$）等于 $\exp(-\alpha|\omega|)$。因此，虽然柯西随机变量的矩不存在或是无限的，但它有特征函数：

$$\Phi_X(\omega)=\exp(\mathrm{j}\omega c-\alpha|\omega|) \tag{5-105}$$

对于均值：

$$\begin{aligned}\mathcal{E}[X]&=\frac{\mathrm{d}}{\mathrm{d}(\mathrm{j}\omega)}\exp(\mathrm{j}\omega c-\alpha|\omega|)\Big|_{\omega=0}\\&=(\mathrm{j}\omega c-\alpha|\omega|)\exp(\mathrm{j}\omega c-\alpha|\omega|)\left(c-\alpha\frac{\mathrm{d}|\mathrm{j}\omega|}{\mathrm{d}(\mathrm{j}\omega)}\right)\Big|_{\omega=0}\end{aligned} \tag{5-106}$$

其中代入了 $|\omega|=|\mathrm{j}\omega|$。由于没有定义 $\omega=0$ 处的绝对值的导数，因此均值 $\mathcal{E}[X]$ 不存在。

我们在随机向量的多元特征函数的定义中结束本节。◀

定义（多元特征函数）　**多元特征函数**是下列联合期望：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})\triangleq\mathcal{E}_{\boldsymbol{X}}[\exp(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{X})] \tag{5-107}$$

其中 $\boldsymbol{X}=[X_1,\cdots,X_N]^{\mathrm{T}}$，$\boldsymbol{\omega}=[\omega_1,\cdots,\omega_N]^{\mathrm{T}}$ 期望要利用 $\boldsymbol{X}$ 的联合概率密度函数。

例 5-16 均值向量为 $\boldsymbol{\mu}_X$，协方差矩阵为 $\boldsymbol{C}_{XX}$ 的多元高斯随机变量的特征函数为：

$$\begin{aligned}\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=&\frac{1}{(2\pi)^{N/2}|\boldsymbol{C}_{XX}|^{1/2}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp(-(1/2)(\boldsymbol{x}-\boldsymbol{\mu}_X)^{\mathrm{T}}\boldsymbol{C}_{XX}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_X))\\&\cdot\exp(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x})\mathrm{d}x_1\cdots\mathrm{d}x_N\end{aligned} \tag{5-108}$$

可以看出 $\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})$ 的形式类似附录 A 中单个高斯随机变量的特征函数（见习题 5-42）。它由下式给出：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=\exp(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\mu}_X-(1/2)\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{C}_{XX}\boldsymbol{\omega}) \tag{5-109}$$ ◀

例 5-17 考虑和式 $Z=X+Y$，其中 X 和 Y 是独立随机变量，则 Z 的特征函数为：

$$\begin{aligned}\Phi_Z(\omega)&=\mathcal{E}_{XY}[\exp(\mathrm{j}\omega(X+Y))]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(\mathrm{j}\omega x)\exp(\mathrm{j}\omega y)f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y\end{aligned} \tag{5-110}$$

由于 X 和 Y 是独立的，$f_{X,Y}(x,y)=f_X(x)f_Y(y)$，因此：

$$\Phi_Z(\omega)=\mathcal{E}_X[\exp(\mathrm{j}\omega X)]\mathcal{E}_Y[\exp(\mathrm{j}\omega Y)]=\Phi_X(\omega)\Phi_Y(\omega) \tag{5-111}$$

利用傅里叶变换的性质，我们知道频域相乘对应时域卷积。由于特征函数的逆变换是概率密度函数，我们得到 $f_Z(z)=f_X(x)*f_Y(y)$，这与第 4 章中利用变换方法得到的结果相同。习题 5-18 考虑了随机变量的差 $Z=X-Y$。◀

5.12　条件期望

第 2 章介绍了抽象概率空间 $\{\Omega,\mathcal{F},\mathcal{P}\}$ 中事件 A 和 B 的条件概率，讨论了以事件 B 为条件是如何简化事件 A 的样本空间。即

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{5-112}$$

假设 $P(B)\neq0$。在第 3 章，我们认为条件概率密度函数的形式类似式(5-112)：

$$f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)},\quad f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)} \tag{5-113}$$

条件概率质量函数也是如此。条件期望，也就是所谓的条件均值，其形式与上述表达式类似。

定义(条件期望) 给定随机变量$Y=y$时，随机变量X的**条件期望**为：

$$\mathcal{E}[X|Y=y]\triangleq\int_{-\infty}^{\infty}xf_{X|Y}(x|y)\mathrm{d}x \tag{5-114}$$

由于$\mathcal{E}[X|Y=y]$是一个y的非随机函数，因此可以写成$\mathcal{E}[X|Y=y]=g(y)$。随机变量Y的条件期望本身是一个随机变量，因为它是Y的函数：$\mathcal{E}[X|Y]=g(y)$。可以先将$\mathcal{E}[X|Y=y]=g(y)$看成y的函数，然后将y替换成Y，得到$\mathcal{E}[X|Y]$。结果$Z\triangleq g(Y)$是一个随机变量，其分布$f_Z(z)$通过对$f_Y(y)$进行适当的变换得到。考虑一个能说明两种条件期望的区别的一个简单例子。

例 5-18 令$Y=X+V$，其中X是对称的伯努利随机变量，等概率取±1，V是参数$\{\mu=0,\sigma\}$的高斯分布。例 4-15 中给出了Y的概率密度函数，以及给定$X=\pm1$时Y的条件概率密度函数。利用$X=1$时的条件概率密度函数得到：

$$\begin{aligned}\mathcal{E}[Y|X=1]&=\int_{-\infty}^{\infty}yf_{Y|X}(y|1)\mathrm{d}y\\&=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}y\exp(-(y-1)^2/2\sigma^2)\mathrm{d}y=1\end{aligned} \tag{5-115}$$

因为$X=1$时Y的条件概率密度函数是均值为 1 的高斯分布，很容易得到上述结果。类似的，$X=-1$时Y的条件概率密度函数是均值为-1的高斯分布。接下来将以随机变量X为条件。根据前面的结果，很容易得出$\mathcal{E}[X|Y=x]=x$，从而

$$\mathcal{E}[Y|X]=X \tag{5-116}$$

现在假设V的均值是μ_V。然后直接得出$\mathcal{E}[Y|X=1]=\mu_V+1$，$\mathcal{E}[Y|X=-1]=\mu_V-1$，从而：

$$\mathcal{E}[Y|X]=\mu_V+X \tag{5-117}$$ ◀

例 5-19 考虑以下随机变量的和：

$$Y=\sum_{m=1}^{N}X_m \tag{5-118}$$

其中$\{X_m\}$是独立同分布(iid)的伯努利随机变量，每个参数都为p。假设N是分布未知的正的离散随机变量。如果固定$N=n$，那么Y的概率密度函数是参数为$\{n,p\}$的二项分布。从而：

$$\mathcal{E}[Y|N=n]=np \tag{5-119}$$

它是一个常数，而：

$$\mathcal{E}[Y|N]=Np \tag{5-120}$$

它是一个随机变量。 ◀

例 5-20 假设利用第二个随机变量X的函数估计随机变量Y。该方法与第 9 章介绍的估计理论有关。用$\hat{Y}$表示Y的估计，并假定$\hat{Y}=g(X)$。性能指标(代价函数)采用均方误差(MSE)，其期望利用了X和Y的联合概率密度：

$$\begin{aligned}\mathcal{E}[(Y-g(X))^2]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-g(x)]^2f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-g(x)]^2f_{Y|X}(y|x)f_X(x)\mathrm{d}x\mathrm{d}y\end{aligned} \tag{5-121}$$

将式(5-121)关于$g(y)$进行微分，得到：

$$\frac{\partial}{\partial g(x)}\mathcal{E}[(Y-g(X))^2]=-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-g(x)]f_{Y|X}(y|x)f_X(x)\mathrm{d}x\mathrm{d}y \tag{5-122}$$

令结果等于 0 得到：

$$\begin{aligned}\int_{-\infty}^{\infty}f_X(x)\int_{-\infty}^{\infty}yf_{Y|X}(y|x)\mathrm{d}y\mathrm{d}x&=\int_{-\infty}^{\infty}f_X(x)g(x)\int_{-\infty}^{\infty}f_{Y|X}(y|x)\mathrm{d}y\mathrm{d}x\\&=\int_{-\infty}^{\infty}f_X(x)g(x)\mathrm{d}x\end{aligned} \tag{5-123}$$

其中等式右边第二个关于 y 的积分等于1。该等式对所有 x 成立，而且由于 $f_X(x)\geqslant 0$，得到：

$$g(x)=\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\mathrm{d}y=\mathcal{E}[Y|X=x] \tag{5-124}$$

最后，由于 X 是一个随机变量，可以写成：

$$\hat{Y}=g(X)=\mathcal{E}[Y|X] \tag{5-125}$$

这也是一个随机变量。给定 X 时 Y 的最小均方误差(MMSE)估计称为条件均值估计器。◀

5.13 条件期望的性质

接下来，我们给出条件期望的几个性质，其中许多性质是普通期望的直接推广。

- 线性　对于常数 a，$b\in\mathcal{R}$，

$$\mathcal{E}[aX_1+bX_2|Y]=a\,\mathcal{E}[X_1|Y]+b\,\mathcal{E}[X_2|Y] \tag{5-126}$$

从期望的定义很容易得到该结果。

- 独立性　对于独立的 X 和 Y，

$$\mathcal{E}[Y|X]=\mathcal{E}[Y] \tag{5-127}$$

由于对于每个 x 都有 $f_{Y|X}(y|x)=f_Y(y)$，条件期望可以简化为：

$$\mathcal{E}[Y|X=x]=\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\mathrm{d}y=\int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y=\mathcal{E}[Y] \tag{5-128}$$

- 条件变量的函数　对于 Y 的函数 $g(Y)$，有以下三个结果：

$$\mathcal{E}[g(Y)|Y]=g(Y) \tag{5-129}$$

$$\mathcal{E}[Xg(Y)|Y]=g(Y)\,\mathcal{E}[X|Y] \tag{5-130}$$

$$\mathcal{E}[X|Y,g(Y)]=\mathcal{E}[X|Y] \tag{5-131}$$

式(5-129)直觉上容易被接受，但要首先通过下列条件累积分布函数进行推导：

$$\begin{aligned}P(Y\leqslant y|Y=y_0)&=\begin{cases}1, & y\geqslant y_0\\ 0, & y<y_0\end{cases}\\ &=u(y-y_0)\end{aligned} \tag{5-132}$$

其中 $u(y)$ 是单位阶跃函数。该结果是有效的，且边缘概率密度函数 $f_Y(y)$ 无关。关于 y 进行微分得到移位的狄拉克 δ 函数 $\delta(y-y_0)$，并且

$$\mathcal{E}[g(Y)|Y=y_0]=\int_{-\infty}^{\infty} g(y)\delta(y-y_0)\mathrm{d}y=g(y_0) \tag{5-133}$$

由于式(5-133)对所有 Y 值都成立，我们得到式(5-129)的结果。类似式(5-130)的结果

$$\begin{aligned}P(X\leqslant x,Y\leqslant y|Y=y_0)&=\begin{cases}P(X\leqslant x|Y=y_0), & y\geqslant y_0\\ 0, & y<y_0\end{cases}\\ &=P(X\leqslant x|Y=y_0)u(y-y_0)\end{aligned} \tag{5-134}$$

对 x 和 y 微分得到 $f_{X|Y}(x|y_0)\delta(y-y_0)$，因此：

$$\begin{aligned}\mathcal{E}[Xg(Y)|Y=y_0]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xg(y)f_{X|Y}(x|y_0)\delta(y-y_0)\mathrm{d}x\mathrm{d}y\\ &=\int_{-\infty}^{\infty} xf_{X|Y}(x|y_0)\mathrm{d}x\int_{-\infty}^{\infty} g(y)\delta(y-y_0)\mathrm{d}y\\ &=g(y_0)\,\mathcal{E}[X|Y=y_0]\end{aligned} \tag{5-135}$$

该结果适用于所有的 Y 值，式(5-130)的结果由条件期望的因子分解性质得证。式(5-131)的性质里，$g(Y)$ 未能提供超出 Y 本身能够给出的更多信息；事实上，如果 $g(\cdot)$ 不是一对一的，它包含的信息少于 Y。因此累积分布函数与式(5-134)相同：

$$P(X\leqslant x|Y=y_0,g(Y)=g(y_0))=P(X\leqslant x|Y=y_0)u(y-y_0) \tag{5-136}$$

使得：

$$\mathcal{E}[X|Y=y_0,g(Y)=g(y_0)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf_{X|Y}(x|y_0)\delta(y-y_0)\mathrm{d}x\mathrm{d}y$$

$$=\mathcal{E}[X|Y=y_0] \tag{5-137}$$

式(5-131)得证。

- **两个变量的嵌套期望**　对于随机变量 X 和 Y，

$$\mathcal{E}[\mathcal{E}[X|Y]]=\mathcal{E}[X] \tag{5-138}$$

外期望是关于 Y 的：

$$\begin{aligned}\mathcal{E}[\mathcal{E}[X|Y]]&=\int_{-\infty}^{\infty} f_Y(y)\,\mathcal{E}[X|Y=y]\mathrm{d}y=\int_{-\infty}^{\infty} f_Y(y)\int_{-\infty}^{\infty} xf_{X|Y}(x|y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{\infty} xf_X(x)\mathrm{d}x=\mathcal{E}[X]\end{aligned} \tag{5-139}$$

其中利用了式(5-113)。该证明表明，像式(5-139)第一行写的那样，随机变量 $\mathcal{E}[X|Y]$ 可以写成只关于特定值 $Y=y$ 的积分。

例 5-21　对例 5-19 中给定的条件，从式(5-120)得出：

$$\mathcal{E}[Y]=\mathcal{E}_N[\mathcal{E}[Y|N]]=\mathcal{E}[Np]=\mathcal{E}[N]p \tag{5-140}$$

其中 $p=\mathcal{E}[X_m]$。利用嵌套期望，找到期望 $\mathcal{E}[Y]$ 是很简单的，它是独立随机变量的非平凡函数。由于我们感兴趣的只是随机变量 Y 的期望，因此没有必要计算 $\{X_m\}$ 和 N 的联合概率。◀

- **正交性**　对于随机变量 X、Y 和任意函数 $g(\cdot)$：

$$\mathcal{E}[(X-\mathcal{E}[X|Y])g(Y)]=0 \tag{5-141}$$

在该式中，$\mathcal{E}[X|Y]$ 可以看作给定观测值 Y 时 X 的估计(前面提到过)，从而 $E\triangleq X-\mathcal{E}[X|Y]$ 是一个"误差"随机变量。

上述正交原理表明，该误差和 Y 的任意函数是正交的：$\mathcal{E}[Eg(Y)]=0$。随机变量的正交性在本章稍后将再次讨论。重新排列表达式得到：

$$\mathcal{E}[Xg(Y)]=\mathcal{E}[\mathcal{E}[Xg(Y)|Y]] \tag{5-142}$$

这里我们利用了式(5-130)的因子分解性质(反过来用)。考虑 $Y=y$ 的内期望：

$$\mathcal{E}[Xg(Y)|Y=y]=\int_{-\infty}^{\infty} xg(y)f_{X|Y}(x|y)\mathrm{d}x \tag{5-143}$$

式(5-142)右侧的外期望为：

$$\int_{-\infty}^{\infty} f_Y(y)\int_{-\infty}^{\infty} xg(y)f_{X|Y}(x|y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xg(y)f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{5-144}$$

这是式(5-142)的左边。从式(5-144)观察到式(5-142)右侧的内期望是关于 $X|Y=y$ 的，外期望是关于 Y 的，而左侧的期望是关于 X 和 Y 的联合概率密度的。因此，需要注意如何理解和计算期望，如式(5-141)的外期望就涉及多项的组合。

考虑以下期望：

$$\mathcal{E}[g(Y)a(X,Y)-g(Y)b(Y)]=0 \tag{5-145}$$

该式类似于式(5-141)，但函数 $a(X, Y)$ 取代了 X，函数 $b(Y)$ 取代了 $\mathcal{E}[X|Y]$。它表明如果式(5-145)中的条件适用于任何函数 $g(Y)$，则 $b(Y)$ 一定是第一项的条件期望乘以 $g(Y)$，也就是 $b(Y)=\mathcal{E}[a(X,Y)|Y]$。这是一个有用的结论，因为它可以得出期望的等价计算，具体方法是将式(5-145)按以下方式重新排列(Larson 和 Shubert，1979)：

$$\mathcal{E}[g(Y)a(X,Y)]=\mathcal{E}[g(Y)b(Y)] \tag{5-146}$$

一般情况下，尽管该期望等式适用于任何 $g(Y)$ 函数，但不能得出 $a(X,Y)$ 和 $b(Y)$ 相等；只能说 $b(Y)=\mathcal{E}[a(X,Y)|Y]$ 是条件期望。然而，如果 $a(X,Y)\to a(Y)$ 只是 Y 的函数，那么由式(5-129)的性质可以得出 $b(Y)=\mathcal{E}[a(Y)|Y]=a(Y)$。

例 5-22　这里举例说明如何利用上述结论再次证明因子分解性质。通过将式(5-141)中的 $g(Y)$ 乘以 $h(Y)$，我们可以将式(5-130)的左侧写成以下期望：

$$\mathcal{E}[Xg(Y)h(Y)]=\mathcal{E}[\mathcal{E}[Xg(Y)|Y]h(Y)] \tag{5-147}$$

该结论适用于任何 $h(Y)$。接下来，定义 $c(Y)\triangleq g(Y)h(y)$，它同样是 Y 的函数。利用

式(5-141)得出，对于任意 $c(Y)=g(Y)h(Y)$：

$$\mathcal{E}[Xc(Y)]=\mathcal{E}[\mathcal{E}[X|Y]c(Y)]=\mathcal{E}[\mathcal{E}[X|Y]g(Y)h(Y)] \tag{5-148}$$

由于 $g(Y)$ 是给定的，上式也适用于任意 $h(Y)$。式(5-147)与式(5-148)的左侧相同，通过右侧得到：

$$\mathcal{E}[\mathcal{E}[Xg(Y)|Y]h(Y)]=\mathcal{E}[\mathcal{E}[X|Y]g(Y)h(Y)] \tag{5-149}$$

适用于任意 $h(Y)$。因此，从上面的讨论可知，期望中乘以 $h(Y)$ 的两部分相等，即：$\mathcal{E}[Xg(Y)|Y]=\mathcal{E}[X|Y]g(Y)$，这是因子分解性质。◀

正交原理是 MSE 估计理论中使用的一个重要性质，第 11 和 12 章将介绍更多正交原理的细节。

● **三个变量的嵌套期望**　对于随机变量 X、Y 和 Z，

$$\mathcal{E}_{Y|Z}[\mathcal{E}_{X|Y,Z}[X|Y,Z]|Z]=\mathcal{E}_{X|Z}[X|Z] \tag{5-150}$$

正如之前所提到的，需要写出特定 z 值的条件期望：

$$\begin{aligned}\mathcal{E}_{Y|Z}[\mathcal{E}_{X|Y,Z}[X|Y,Z]|Z=z]&=\int_{-\infty}^{\infty}\mathcal{E}[X|Y=y,Z=z]f_{Y|Z}(y|z)\mathrm{d}y\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{X|Y,Z}(x|y,z)f_{Y|Z}(y|z)\mathrm{d}y\mathrm{d}x\end{aligned} \tag{5-151}$$

重写每个条件概率密度函数，得到

$$\mathcal{E}_{Y|Z}[\mathcal{E}_{X|Y,Z}[X|Y,Z]|Z=z]=\frac{1}{f_Z(z)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}f_{Y,Z}(y,z)\mathrm{d}y\mathrm{d}x \tag{5-152}$$

两个概率密度函数相互抵消，Y 从剩下的联合概率密度函数中积分出来，得到：

$$\begin{aligned}\mathcal{E}_{Y|Z}[\mathcal{E}_{X|Y,Z}[X|Y,Z]|Z=z]&=\frac{1}{f_Z(z)}\int_{-\infty}^{\infty}xf_{X,Z}(x,z)\mathrm{d}x\\&=\int_{-\infty}^{\infty}xf_{X|Z}(x|z)\mathrm{d}x=\mathcal{E}_{X|Z}[X|Z=z]\end{aligned} \tag{5-153}$$

它适用于所有的 z，因此得到式(5-150)的右边表达式。

先前的结果一般适用于三个随机变量，通过适当的条件嵌套可以用于解决复杂期望。在学习例 5-23 提供的部分背景知识后，我们证明了同样涉及三个随机变量的定理，但要求三个变量有顺序的粗糙 σ 域以得出有意思的性质。

例 5-23　考虑离散样本空间 Ω 并定义两个 σ 域，使得：

$$\mathcal{F}_2\subset\mathcal{F}_1 \tag{5-154}$$

这意味着 $\mathcal{F}_2$ 比 $\mathcal{F}_1$ 粗略；同时基数 $|\mathcal{F}_2|<|\mathcal{F}_1|$。用 $\{E_n\}$ 和 $\{F_m\}$ 分别表示 $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 的元素(见第 2 章元素的讨论)。图 5-7 给出了掷一个骰子的元素的例子。其中 $\mathcal{F}_1$ 是 $\mathcal{P}(\Omega)$ 的幂集且 $|\mathcal{F}_1|=64$，$|\mathcal{F}_2|=\{\phi, \Omega$，偶数，奇数$\}$ 且 $|\mathcal{F}_2|=4$。显然元素 $\{F_m\}$ 没有元素 $\{E_n\}$ 精细，因此这个例子满足式(5-154)。定理 5-6 将式(5-154)扩展到三个 σ 域，以展示条件期望的平滑性质。

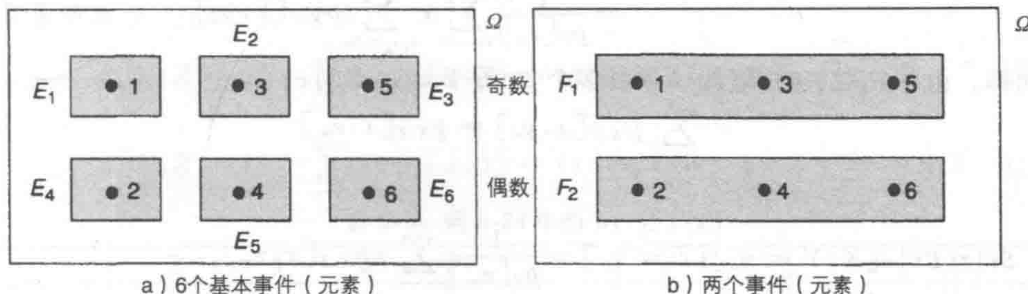


图 5-7　例 5-23 中掷一个骰子的两个 σ 域的元素。a) 事件 $\{E_n\}$ 有 6 个结果，通过幂集 $P(\Omega)$ 得到精细 σ 域；b) 事件 $\{F_m\}$ 有奇数和偶数的元素，通过 $\{\phi, \Omega$，偶数，奇数$\}$ 得到粗糙 σ 域 ◀

考虑结果 $x \in \mathcal{Z}$ 的离散随机变量 X，x 是 σ 域 $\mathcal{F}_X = \mathcal{P}(Z)$（这些结果的幂集）的元素。如果另一个离散随机变量 Y 的 σ 域 $\mathcal{F}_Y$ 是 $\mathcal{F}_X$ 的一个子集，那么这意味着 $\mathcal{F}_Y$ 的元素（结果 y）是 $\mathcal{F}_X$ 的元素（结果 x）的子集。很容易想象 Y 和 X 的多对一映射：一个特定的 y 可以通过多个 x 获得，而特定 x 只对应一个 y。图 5-7 显然是这种情况，$x=\{1,2,3,4,5,6\}$，$y=\{$偶数，奇数$\}$，$\{2,4,6\} \rightarrow$ 偶数，$\{1,3,5\} \rightarrow$ 奇数。由于讨论的是随机变量，我们可以使用 $\{0,1\}$ 代替{偶数，奇数}。

定理 5-6（平滑性） 设 X、Y 和 Z 是 $\mathcal{Z}$ 上的离散随机变量，三个不等的 σ 域满足

$$\mathcal{F}_Z \subset \mathcal{F}_Y \subset \mathcal{F}_X \tag{5-155}$$

依次取条件 Y 和 Z，与仅取条件 Z 相等：

$$\mathcal{E}_{Z|Y}[\mathcal{E}_{X|Y}[X|Y]|Z] = \mathcal{E}_{X|Z}[X|Z] \tag{5-156}$$

反转条件次序后，

$$\mathcal{E}_{Y|Z}[\mathcal{E}_{X|Z}[X|Z]|Y] = \mathcal{E}_{X|Z}[X|Z] \tag{5-157}$$

观察发现式(5-156)、式(5-157)中的条件期望是不对称的，交换条件次序后，结果仍然是一个粗糙 σ 域随机变量的条件期望。利用多对一函数 $Y=g(X)$ 和 $Z=h(Y)$ 可以产生满足式(5-155)的随机变量。因此，X 的 σ 域必然是三个随机变量中最精细的。由于是多对一映射，不能将 X 写成 Y 和 Z 的函数。

证明：（离散随机变量） 令 Y 和 Z 的元素分别是 $\{y_n\}$，$n=1, \cdots, |\mathcal{F}_Y|$ 和 $\{z_m\}$，$m=1, \cdots, |\mathcal{F}_Z|$。因为 $\mathcal{F}_Z$ 比 $\mathcal{F}_Y$ 粗糙，证明的关键是要认识到：

$$\begin{aligned} p_{Y,Z}[y_n, z_m] &= \begin{cases} p_Y[y_n], & y_n = z_m \\ 0, & \text{其他} \end{cases} \\ &= p_Y[y_n] I_{\{z_m\}}[y_n] \end{aligned} \tag{5-158}$$

其中 $I_{\{z_m\}}[y_n]$ 是离散指标函数。由于 y_n 只与 Z 的结果有关，因此只有当精细 σ 域的结果 y_n 与粗糙 σ 域的特定值 z_m 相同时，该概率非零。因此：

$$p_{Y|Z}[y_n|z_m] = \frac{p_{Y,Z}[y_n, z_m]}{p_Z[z_m]} = \frac{p_Y[y_n]}{p_Z[z_m]} I_{\{z_m\}}[y_n] \tag{5-159}$$

式(5-156)的外期望与 $\{y_n\}$ 的所有结果有关，因此：

$$\begin{aligned} \mathcal{E}_{Y|Z}[\mathcal{E}_{X|Y}[X|y_n]|z_m] &= \sum_{n=1}^{|\mathcal{F}_Y|} \mathcal{E}_{X|Y}[X|y_n] p_{Y|Z}[y_n|z_m] \\ &= \sum_{n:y_n=z_m} \frac{p_Y[y_n]}{p_Z[z_m]} \mathcal{E}_{X|Y}[X|y_n] \end{aligned} \tag{5-160}$$

假设 $x \in \mathcal{Z}$，内期望为：

$$\begin{aligned} \mathcal{E}_{Y|Z}[\mathcal{E}_{X|Y}[X|y_n]|z_m] &= \sum_{n:y_n=z_m} \frac{1}{p_Z[z_m]} \sum_{x=-\infty}^{\infty} x p_{X|Y}[x|y_n] p_Y[y_n] \\ &= \frac{1}{p_Z[z_m]} \sum_{x=-\infty}^{\infty} x \sum_{n:y_n=Zm} p_{X,Y}[x, y_n] \end{aligned} \tag{5-161}$$

同样，由于 $\mathcal{F}_Z \subset \mathcal{F}_Y$，因此一个或多个 y_n 等于 z_n，得到：

$$\sum_{n:y_n=z_m} p_{X,Y}[x, y_n] = p_{X,Z}[x, z_m] \tag{5-162}$$

使得：

$$\begin{aligned} \mathcal{E}_{Y|Z}[\mathcal{E}_{X|Y}[X|Y=y_n]|Z=z_m] &= \frac{1}{p_Z[z_m]} \sum_{x=-\infty}^{\infty} x p_{X,Z}[x, z_m] \\ &= \sum_{x=-\infty}^{\infty} x p_{X|Z}[x|z_m] = \mathcal{E}_{X|Z}[X|Z=z_m] \end{aligned} \tag{5-163}$$

由于期望适用于所有 $\{y_n\}$ 和 $\{z_m\}$，所以可以将它们替换为随机变量 Y 和 Z，从而完成了证明。

对于式(5-157)中的条件期望，观察式(5-158)发现：

$$p_{Z|Y}[z_m|y_n]=\frac{p_{Z,Y}[z_m,y_n]}{p_Y[y_n]}=\frac{p_Y[y_n]}{p_Y[y_n]}I_{\{z_m\}}[y_n]=I_{\{z_m\}}[y_n] \tag{5-164}$$

从而：

$$\begin{aligned}\mathcal{E}_{Z|Y}[\mathcal{E}_{X|Z}[X|z_m]|y_n]&=\sum_{m=1}^{|\mathcal{F}_Z|}\mathcal{E}_{X|Z}[X|z_m]p_{Z|Y}[z_m|y_n]\\&=\sum_{m:y_n=zm}\mathcal{E}_{X|Z}[X|z_m]=\mathcal{E}_{X|Z}[X|z_m]\end{aligned} \tag{5-165}$$

由于 z_m(粗糙$\mathcal{F}_Z$)只有一个元素与 y_n 相等，因此最后一个表达式的求和中只有一项非0。通过用 Y 和 Z 取代 y_n 和 z_m，得到了最后的结果。

对包含粗糙条件元素的精细元素求平均，并不改变原来的条件期望，因此式(5-156)称为条件期望的平滑性质。也可以说粗糙条件对精细条件有“平滑”作用。

例 5-24 这个例子演示了如何通过将随机变量 X 映射到其他随机变量，来建立粗糙 σ 域。考虑掷骰子试验，其中随机变量 X 的结果是{1, 2, 3, 4, 5, 6}，每个结果的概率是1/6。最大的 σ 域是$\mathcal{P}(\Omega)$的幂集，包含 2^6 个元素。定义 X 到 Y 的映射：

$$Y=\begin{cases}0, & X=1,2\\-1, & X=3,4\\1, & X=5,6\end{cases} \tag{5-166}$$

Y 的 σ 域有 $2^3=8$ 个元素，显然比较粗糙：$\mathcal{F}_Y\subset\mathcal{F}_X$。注意随机变量 Y 有两种观察角度：(ⅰ)与 X 的样本空间相同，但 σ 域$\mathcal{F}_Y$ 比较粗糙；(ⅱ)样本空间缩小为 $\Omega_Y=\{0,-1,1\}$，σ 域通过 Ω_Y 的幂集得到。我们感兴趣的是第一种观点：Y 具有粗糙 σ 域。

接下来，定义从 Y 到 Z 的映射：

$$Z=|Y|=\begin{cases}0, & Y=0\\1, & Y=-1,1\end{cases} \tag{5-167}$$

其中$\mathcal{F}_Z\subset\mathcal{F}_Y$ 且只有 $2^2=4$ 个元素：$\{\phi,\Omega,0,1\}$。图 5-8 给出了从 X 到 Y，以及从 Y 到 Z 的多对一映射。通过分析得到以下概率密度函数：

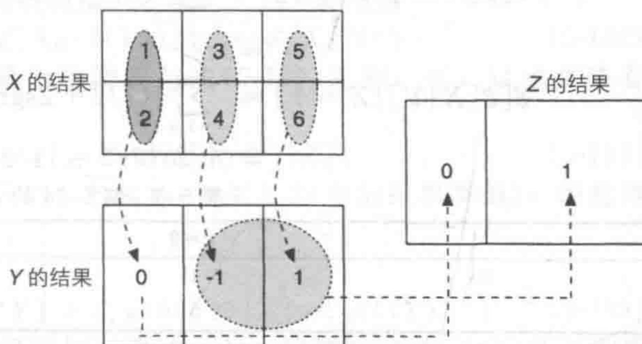


图 5-8　通过映射 $X\to Y$ 和 $Y\to Z$，得到例 5-24 中越来越粗糙的 σ 域

$$f_X(x)=(1/6)\sum_{n=1}^{6}\delta(x-n) \tag{5-168}$$

$$f_Y(y)=(1/3)[\delta(y+1)+\delta(y)+\delta(y-1)] \tag{5-169}$$

$$f_Z(z)=(1/3)\delta(z)+(2/3)\delta(z-1) \tag{5-170}$$

为了验证条件期望的平滑性质，首先计算$\mathcal{E}[X|Z]$，这就会用到表 5-1 中总结的条件概率质量函数 $p_{X|Z}[x|z]$。从而：

$$\mathcal{E}[X|Z=0]=\sum_{x=1}^{6}xp_{X|Z}[x|0]=(1/2)\times(1+2)=3/2 \tag{5-171}$$

$$\mathcal{E}[X|Z=1)=\sum_{x=1}^{6}xp_{X|Z}[x|1]=(1/4)\times(3+4+5+6)=9/2 \tag{5-172}$$

表 5-1　例 5-24 中的 $p_{X|Z}[x|z]$

	X=1	2	3	4	5	6
z= 0	1/2	1/2	0	0	0	0
1	0	0	1/4	1/4	1/4	1/4

由于$\mathcal{E}[X|Z]$是随机变量 Z 的函数，根据以上结果得到：

$$\mathcal{E}[X|Z]=3/2+3Z \tag{5-173}$$

接下来计算$\mathcal{E}[\mathcal{E}[X|Y]|Z]$以表明随机变量 Y(有更精细的 σ 域)与$\mathcal{E}[X|Z]$具有相同的结果。表 5-2 总结了条件概率质量函数 $p_{X|Z}[x|y]$，因此内期望：

$$\mathcal{E}[X|Y=0]=\sum_{x=1}^{6}xp_{X|Y}[x|0]=(1/2)(1+2)=3/2 \tag{5-174}$$

$$\mathcal{E}[X|Y=-1]=\sum_{x=1}^{6}xp_{X|Y}[x|-1]=(1/2)(3+4)=7/2 \tag{5-175}$$

$$\mathcal{E}[X|Y=1]=\sum_{x=1}^{6}xp_{X|Y}[x|1]=(1/2)(5+6)=11/2 \tag{5-176}$$

表 5-2 例 5-24 的 $p_{X|Y}[x|y]$

	$X=1$	2	3	4	5	6
0	1/2	1/2	0	0	0	0
$Y=$ −1	0	0	1/2	1/2	0	0
1	0	0	0	0	1/2	1/2

上述结果的一般表达式为：

$$\mathcal{E}[X|Y]=7/2+2\operatorname{sgn}(Y+1)\operatorname{sgn}(Y-1/2) \tag{5-177}$$

其中 sgn(·)为符号函数。利用表 5-3 中的外期望，完成推导

$$\begin{aligned}\mathcal{E}[\mathcal{E}[X|Y]|Z=0]&=\sum_{y=-1,0,1}(7/2+2\operatorname{sgn}(y+1)\operatorname{sgn}(y-1/2))p_{Y|Z}[y|0]\\&=3/2\end{aligned} \tag{5-178}$$

$$\begin{aligned}\mathcal{E}[\mathcal{E}[X|Y]|Z=1]&=\sum_{y=-1,0,1}(7/2+2\operatorname{sgn}(y+1)\operatorname{sgn}(y-1/2))p_{Y|Z}[y|1]\\&=(1/2)(7/2+11/2)=9/2\end{aligned} \tag{5-179}$$

表 5-3 例 5-24 的 $p_{Y|Z}[y|z]$

	$Y=0$	−1	1
$Z=$ 0	1	0	0
1	0	1/2	1/2

这与前面$\mathcal{E}[X|Z]$的结果相同。

表 5-4 总结了条件期望的各种性质。习题 11-14 证明了第三个性质(独立性质 2)，该性质可以用来推导 11 章中的卡尔曼滤波器。通过比较结果，可以得出以下补充结论：

- 嵌套期望 1 通过以下方式得到：重新整理正交性质，然后利用分解性质将 $g(Y)$放到内期望里面。比较嵌套期望 1 和嵌套期望 2 发现，前者可以看作将 X 替换成 $g(Y)X$的嵌套期望的广义形式。
- 平滑性质 1 可以通过将 σ 域依次粗糙的条件代入嵌套期望 3 得到。
- 从技术上看，表中最后一行不是平滑性质，因为内期望的条件也就是随机变量 Y 的 σ 域比 X 粗糙。相反，它与表中第二行的独立性质“外观”上相同，以至于外期望没有产生任何影响。

表 5-4 条件期望的性质

性质	条件斯望	条件
线性	$\mathcal{E}[aX_1+bX_2\mid Y]=a\,\mathcal{E}[X_1\mid Y]+b\,\mathcal{E}[X_2\mid Y]$	a、b 是常数
独立 1	$\mathcal{E}[Y\mid X]=\mathcal{E}[Y]$	X，Y 独立
独立 2	$\mathcal{E}[X\mid Y,Z]=\mathcal{E}[X\mid Y]+\mathcal{E}[X\mid Z]-\mathcal{E}[X]$	Y，Z 独立，X，Y，Z 服从联合高斯分布
Y 的函数	$\mathcal{E}[g(Y)\mid Y]=g(Y)$	无
Y 的函数	$\mathcal{E}[X\mid Y,g(Y)]=\mathcal{E}[X\mid Y]$	无
因子分解	$\mathcal{E}[Xg(Y)\mid Y]=g(Y)\mathcal{E}[X\mid Y]$	无

（续）

性质	条件斯望	条件
正交	$\mathcal{E}[(X-\mathcal{E}[X\|Y])g(Y)]=0$	无
嵌套1	$\mathcal{E}[\mathcal{E}[Xg(Y)\|Y]]=\mathcal{E}(Xg(Y)]$	无
嵌套2	$\mathcal{E}[\mathcal{E}[X\|Y]]=\mathcal{E}[X]$	无
嵌套3	$\mathcal{E}[\mathcal{E}[X\|Y,Z]\|Z]=\mathcal{E}[X\|Z]$	无
平滑	$\mathcal{E}[\mathcal{E}[X\|Y]\|Z]=\mathcal{E}[X\|Z]$	$\mathcal{F}_Z\subset\mathcal{F}_Y\subset\mathcal{F}_X$
"平滑"	$\mathcal{E}[\mathcal{E}[X\|Z]\|Y]=\mathcal{E}[X\|Z]$	$\mathcal{F}_Z\subset\mathcal{F}_Y\subset\mathcal{F}_X$

最后，在正交性质的基础上，给出表5-4中平滑性质的证明(Larson和Shubert，1979)。该性质更简单通用。它可适用于连续和离散随机变量，但粗糙和精细σ域可能没有那么直观。

证明：（任意随机变量）假设随机变量依次是另外变量的函数来满足条件$\mathcal{F}_Z\subset\mathcal{F}_Y\subset\mathcal{F}_X$。首先，利用$X$和$Z$的正交性质得到：

$$\mathcal{E}[Xg(Z)]=\mathcal{E}[\mathcal{E}[X|Z]g(Z)] \tag{5-180}$$

将等式两边的X替换成$\mathcal{E}[X|Z]$，得到：

$$\mathcal{E}[\mathcal{E}[X|Y]g(Z)]=\mathcal{E}[\mathcal{E}[\mathcal{E}[X|Y]g(Z)|Z]] \tag{5-181}$$

这里利用因子分解性质将等式右边的$g(Z)$放到了第二个期望里面。注意如果$X=c(Y)$，也就是X是Y的函数，由表5-4的第四行得到$\mathcal{E}[X=c(Y)|Y]c(Y)=X$，那么这个表达式可以简化为式(5-180)。因此，X不是Y的函数，它具备更精细的σ域。推导中的下一步是令$g(Z)=g(h(Y))$，因此Z具有较粗糙的σ域。上式左侧变成：

$$\mathcal{E}[\mathcal{E}[X|Y]g(h(Y))]=\mathcal{E}[\mathcal{E}[Xg(h(Y))|Y]=\mathcal{E}[Xg(h(Y))] \tag{5-182}$$

这里利用了因子分解性质和正交性质。最后的期望与式(5-180)左侧一样，这也意味着式(5-180)与式(5-181)的右侧是相等的：

$$\mathcal{E}[g(Z)\,\mathcal{E}[X|Z]]=\mathcal{E}[g(Z)\,\mathcal{E}[\mathcal{E}[X|Y]|Z]] \tag{5-183}$$

由于等式适用于所有$g(Z)$，并且都只是Z的函数(与例5-22中的结果类似)，因此得到$\mathcal{E}[X|Z]=\mathcal{E}[\mathcal{E}[X|Z]|Z]$，即平滑性质。

对于表中的最后一行：

$$\mathcal{E}[\mathcal{E}[X|Z]|Y]=\mathcal{E}[g(Z)|Y]=\mathcal{E}[g(h(Y))|Y]=g(h(Y)) \tag{5-184}$$

其中$\mathcal{E}[X|Z]=g(Z)$是Z的函数(不是前面证明用到的$g(\cdot)$)，再次假设$Z=h(Y)$是Y的函数，它有一个较粗糙的σ域。最后结果利用了表中第四行的结论。由前面的定义$g(h(Y))=\mathcal{E}[X|Z]$，命题得证。

5.14 位置参数：均值、中位数和众数

表示概率密度函数中心和位置的参数可以作为随机变量的简单描述，第3章介绍分布性质时也用到了这些参数。本节将介绍这类随机变量的三个参数，尽管对于一些分布三者的取值可能相等。

定义(均值) 随机变量X的**均值**是其期望：

$$\mu_X\triangleq\mathcal{E}[X] \tag{5-185}$$

随机变量的均值可以看作大量试验结果的平均，从函数的观点也可看作奇函数$g(x)=x$对$f_X(x)$加权后的面积，该面积更关注大的x值(正值和负值)。

定义(中位数) 连续随机变量X的**中位数**m_e是累积分布函数的中点：

$$F_X(m_e)=1/2\Rightarrow\int_{-\infty}^{m_e}f_X(x)\mathrm{d}x=\int_{m_e}^{\infty}f_X(x)\mathrm{d}x=1/2 \tag{5-186}$$

相当于$P(X\leqslant m_e)=P(X\geqslant m_e)=1/2$。对于离散随机变量，最后一个表达式改为以下两个条件：

$$P(X\leqslant m_e)\geqslant 1/2 \text{ 和 } P(X\geqslant m_e)\geqslant 1/2 \tag{5-187}$$

与1/2有关的两个不等号的方向相同。代入概率密度函数，得到：

$$\int_{-\infty}^{m_e}\sum_{n=-\infty}^{\infty}P(X=n)\delta(x-n)\mathrm{d}x\geqslant 1/2\Rightarrow\sum_{n=-\infty}^{m_e}p_X[n]\geqslant 1/2 \tag{5-188}$$

$$\int_{m_e}^{\infty}\sum_{n=-\infty}^{\infty}P(X=n)\delta(x-n)\mathrm{d}x\geqslant 1/2\Rightarrow\sum_{n=m_e}^{\infty}p_X[n]\geqslant 1/2 \tag{5-189}$$

由于只计算整数值 n 的概率质量，为保证两个求和都包括 m_e 并且都大于 1/2，显然中位数必须是整数(对应 X 的一个结果)。该定义的示例如下。

例 5-25 考虑不能写出中位数 m_e 的确切表达式的泊松分布(与附录 A 中的多数随机变量不同)。设 $\alpha=2.5$，由于：

$$P(X\leqslant 2)=\exp(-2.5)[1+2.5/1!+(2.5)^2/2!]\approx 0.5438>1/2 \tag{5-190}$$

从而

$$P(X>2)\approx 1-0.5438=0.4562 \tag{5-191}$$

对于式(5-187)的第二个不等式：

$$P(X\geqslant 2)\approx 0.4562+\exp(-2.5)(2.5)^2/2!\approx 0.7127>1/2 \tag{5-192}$$

从而得出该 α 值对应的 $m_e=2$。◀

虽然离散随机变量的中位数有时取两个结果的中值，对应到例 5-25 的 X 值为 2.5，但式(5-187)中定义的强制中位数取随机变量的整数结果。与均值不同，中位数不受“重尾”的影响(先前讨论的柯西分布)，因此它往往很重要。由于中位数不受罕见的极端事件(称为离群)的影响，因此在统计学里有时将它作为中心的测度。

例 5-26 有 $N=b-a+1$ 个结果 $\{a,\cdots,b\}$ 的离散均匀分布的随机变量，式(5-189)定义的中位数的两个表达式为：

$$\sum_{n=a}^{m_e}(1/N)\geqslant 1/2\Rightarrow m_e-a+1\geqslant N/2\Rightarrow m_e\geqslant(a+b-1)/2 \tag{5-193}$$

$$\sum_{n=m_e}^{b}(1/N)\geqslant 1/2\Rightarrow b-m_e+1\geqslant N/2\Rightarrow m_e\leqslant(a+b+1)/2 \tag{5-194}$$

由于对常数进行求和，所以比较简单。如果 $a+b$ 为奇数(N 是偶数)，将得到两个不同的 m_e 值，因此从定义来看中位数不存在。如果 $a+b$ 为偶数(N 为奇数)，利用 floor 函数 $m_e=\lfloor(a+b+1)/2\rfloor$ 将得到唯一的 m_e 值。有时取 $m_e=(a+b)/2$，但当 $a+b$ 为奇数时，会得到一个不是 X 结果的非整数值。◀

定义(众数)　当 x 的概率密度函数为全局最大时，x 为随机变量 X 的**众数** m_o。

图 5-9 给出了连续和离散分布的均值、中位数和众数。有多个极大值的概率密度函数将对应多个众数。图 5-10 给出了双峰概率密度函数。

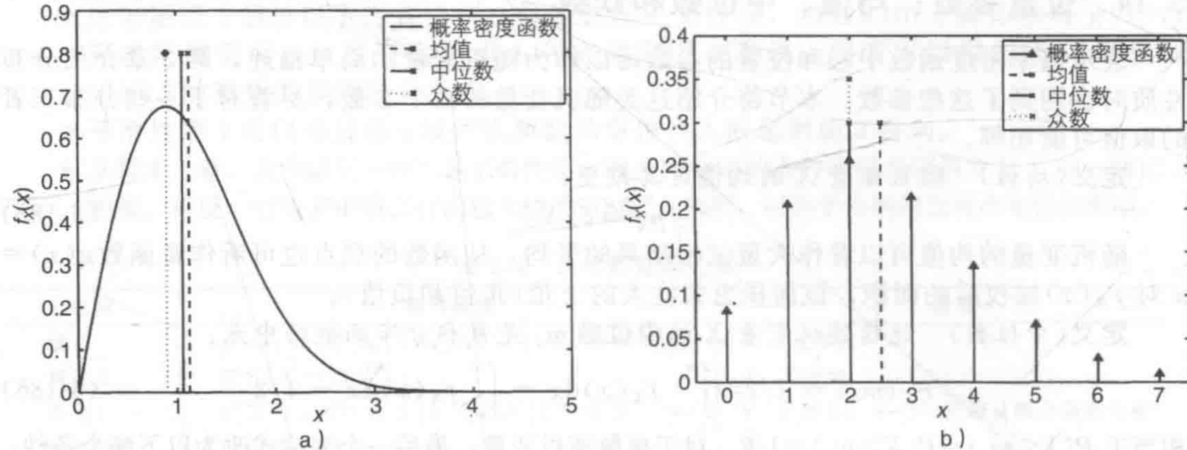


图 5-9　均值、中位数和众数。a) 瑞利分布($\sigma=0.9$)：$\mu_X=\sigma/\sqrt{\pi/2}$，$m_e=\sigma\sqrt{\ln(4)}$，$m_o=\sigma$；b) 泊松分布($\alpha=2.5$)：$\mu_X=\alpha$，$m_e$ 为使用式(5-187)得到的值，$m_o=\lfloor\alpha\rfloor$

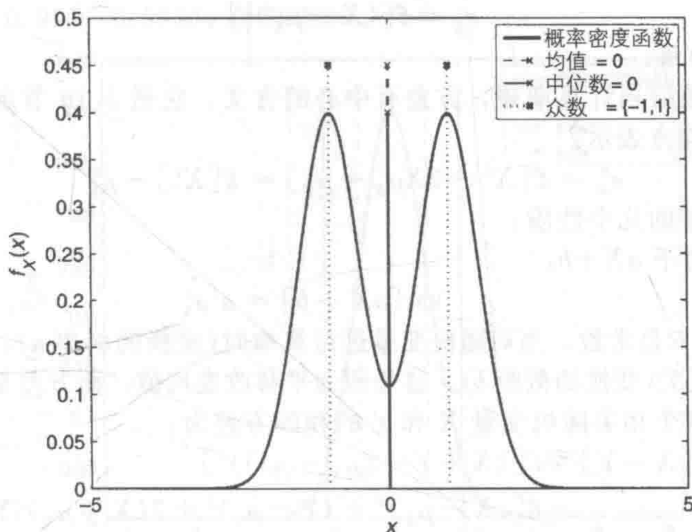


图 5-10　双峰分布示例，该分布通过对两个标准高斯分布的概率密度函数求和得到(每个权值为 0.5)，称为“随机变量的混合”

命题 5-1　随机变量的均值是使均方误差$\mathcal{E}[(X-a)^2]$最小的 a。

证明：按如下方式展开期望：

$$\mathcal{E}[(X-a)^2]=\mathcal{E}[X^2]-2a\,\mathcal{E}[X]+a^2 \tag{5-195}$$

关于 a 进行微分，得到：

$$\frac{\partial}{\partial a}\mathcal{E}[(X-a)^2]=-2\,\mathcal{E}[X]+2a \tag{5-196}$$

令结果等于 0，得到：

$$a=\mathcal{E}[X]=\mu_X \tag{5-197}$$

命题得证。最小均方误差等于方差 σ_X^2。

命题 5-2　随机变量的中位数是使绝对误差的均值$\mathcal{E}[|X-a|]$最小的 a。

证明：按如下方式展开期望：

$$\begin{aligned}\mathcal{E}[|X-a|]&=\int_{-\infty}^{\infty}|x-a|f_X(x)\mathrm{d}x\\&=\int_a^{\infty}(x-a)f_X(x)\mathrm{d}x-\int_{-\infty}^{a}(x-a)f_X(x)\mathrm{d}x\end{aligned} \tag{5-198}$$

利用链式法则关于 a 求微分，并令结果等于 0，得到：

$$-\int_a^{\infty}f_X(x)\mathrm{d}x+\int_{-\infty}^{a}f_X(x)\mathrm{d}x=0 \tag{5-199}$$

式(5-199)与式(5-186)相同，从而 $a=m_e$。

例 5-27　对于参数为$\{\mu,\sigma\}$的高斯随机变量 X，其均值、中位数和众数都等于 μ。参数为 λ 的指数随机变量 X，均值、中位数和众数分别是 $1/\lambda$、$\ln(2)/\lambda$ 和 0。均值通过式(5-49)得到，众数显然是 0。中位数的计算方法如下：

$$\int_0^{m_e}f_X(x)\mathrm{d}x=\lambda\int_0^{m_e}\exp(-\lambda x)\mathrm{d}x=1-\exp(-\lambda m_e)$$

令上式等于 1/2，得到：

$$\exp(-\lambda m_e)=1/2\Rightarrow m_e=\ln(2)/\lambda \tag{5-200}$$ ◀

5.15　方差、协方差和相关

这里将讨论随机变量的分散或扩散程度的测度。

定义(方差)　随机变量 X 的**方差**为：

$$\sigma_X^2 \triangleq \mathcal{E}[(X-\mu_X)^2] \tag{5-201}$$

其中 $\mu_X=\mathcal{E}[X]$ 是均值。

由于是减去均值后再计算期望，方差有中心的含义，它是 5.16 节定义的中心矩之一。方差可以由均值和均方表示：

$$\sigma_X^2 = \mathcal{E}[X^2-2X\mu_X+\mu_X^2] = \mathcal{E}[X^2]-\mu_X^2 \tag{5-202}$$

这里总结了方差的几个性质：

- *线性变换*　对于 $aX+b$：
$$\text{var}[aX+b] = a^2\sigma_X^2 \tag{5-203}$$
其中 $\{a,b\}\in\mathcal{R}$ 是常数。当对随机变量进行展缩时(变换的斜率 a)，方差也随之变化，但是与平移无关(变换的截距 b)。这是因为平移改变均值，而方差是居中的期望。
- *和的方差*　两个相关随机变量 X 和 Y 的和的方差为：
$$\begin{aligned}\text{var}[X+Y] &= \mathcal{E}[(X+Y-(\mu_X+\mu_Y))^2]\\ &= \mathcal{E}[(X-\mu_X)^2+(Y-\mu_Y)^2+2(X-\mu_X)(Y-\mu_Y)]\\ &= \sigma_X^2+\sigma_Y^2+2\text{cov}[X,Y]\end{aligned} \tag{5-204}$$
其中 $\text{cov}[X,Y]\triangleq\mathcal{E}[(X-\mu_X)(Y-\mu_Y)]$ 是 X 和 Y 的协方差 C_{XY}(本节后面将介绍)。如果 X 和 Y 是独立随机变量，则：
$$\text{var}[X+Y] = \sigma_X^2+\sigma_Y^2 \tag{5-205}$$
事实上，该结果只要求 $C_{XY}=0$，也就是随机变量不相关即可。
- *方差的线性组合*　可以按照下列方式将前面的结论扩展到多个随机变量的加权和：
$$\text{var}\left[\sum_{n=1}^{N}a_nX_n\right] = \sum_{n=1}^{N}a_n^2\sigma_{X_n}^2 + \sum_{m=1}^{N}\sum_{n=1,n\neq m}^{N}a_ma_nC_{X_mX_n} \tag{5-206}$$
其中 $\{a_n\}\in\mathcal{R}$ 是常数。
- *差的方差*　两个相关随机变量差的方差是：
$$\text{var}[X-Y] = \sigma_X^2+\sigma_Y^2-2C_{XY} \tag{5-207}$$
如果 X 和 Y 不相关，也就是 $C_{XY}=0$，则 $\text{var}[X-Y]=\sigma_X^2+\sigma_Y^2$ 是各自方差的和；由于方差不能为负，因此它必须是和(而不是差)。

定理 5-7(*总方差法*)　*对于相关随机变量：*

$$\text{var}[X] = \mathcal{E}[\text{var}[X|Y]]+\text{var}[\mathcal{E}[X|Y]] \tag{5-208}$$

证明：取条件 Y 并利用类似式(5-202)的形式得到：

$$\text{var}[X] = \mathcal{E}[\mathcal{E}[X^2|Y]]-(\mathcal{E}[\mathcal{E}[X|Y]])^2 \tag{5-209}$$

其中外期望是关于 Y 的。再次利用式(5-202)并改写第一个期望，得出：

$$\begin{aligned}\text{var}[X] &= \mathcal{E}[\text{var}[X|Y]+(\mathcal{E}[X|Y])^2]-(\mathcal{E}[\mathcal{E}[X|Y]])^2\\ &= \mathcal{E}[\text{var}[X|Y]]+\mathcal{E}[(\mathcal{E}[X|Y])^2]-(\mathcal{E}[\mathcal{E}[X|Y]])^2\end{aligned} \tag{5-210}$$

结合后两项得出另一个方差：

$$\mathcal{E}[(\mathcal{E}[X|Y])^2]-(\mathcal{E}[\mathcal{E}[X|Y]])^2 = \text{var}[\mathcal{E}[X|Y]] \tag{5-211}$$

将该结果代入式(5-210)，定理得证。

该定理与第 2 章的总概率法类似。

定义(*标准差*)　*随机变量 X 的标准差 σ_X 是方差 σ_X^2 的正平方根。*

与方差不同，标准差与均值、随机变量本身有相同的单位，因此通常将标准差作为随机变量分散程度的测度。与方差一样，标准差与随机变量展缩系数有关，与平移系数无关：

$$\text{stdev}[aX+b] = |a|\sigma_X \tag{5-212}$$

其中 a，$b\in\mathcal{R}$ 为常数。

例 5-28　*考虑参数 $\{\mu=0,\ \sigma=1\}$ 的标准高斯随机变量。图 5-11 标出了偏离均值 1 个、2 个、3 个标准差的随机变量的值。当 a 分别等于 1，2，3，4 和 5 时，$P(|X|\leqslant a\sigma)$ 分别*

为 0.683，0.955，0.997，0.9999，0.999999。

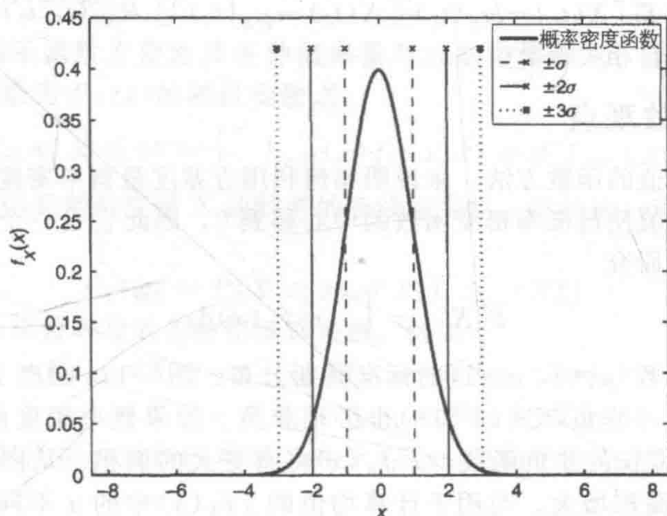


图 5-11 标准高斯概率密度函数，显示了偏离均值 1 个，2 个，3 个标准差的值。这些区间(均值向两侧延伸)对应的概率分别为 0.683，0.955，和 0.997 ◀

定义(变差系数) 随机变量 X 的**变差系数**是：

$$c_X \triangleq \frac{\sigma_X}{\mu_X} \tag{5-213}$$

这里假设 $\mu_X \neq 0$。

它是概率密度函数关于均值差异的测度。因为比值里使用的是标准差而不是方差，其单位与 X 无关。

协方差可以看作一个随机变量的方差到两个随机变量的扩展。

定义(协方差和互相关) 随机变量 X 和 Y 的**协方差**是：

$$C_{XY} \triangleq \mathcal{E}[(X-\mu_X)(Y-\mu_Y)] \tag{5-214}$$

它可以改写为：

$$C_{XY} = \mathcal{E}[XY] - \mu_X\,\mathcal{E}[Y] - \mu_Y\,\mathcal{E}[X] + \mu_X\mu_Y = R_{XY} - \mu_X\mu_Y \tag{5-215}$$

其中

$$R_{XY} \triangleq \mathcal{E}[XY] \tag{5-216}$$

是 X 和 Y 的互相关。

观察发现 $C_{XY}=R_{XY}$要求至少一个随机变量具有零均值。互相关是 X 和 Y 的线性关系的测度，详见后面定义的相关系数。协方差有以下性质：

- 不相关　当 $C_{XY}=0$ 时，随机变量 X 和 Y 不相关，这时 $R_{XY}=\mathcal{E}[X]\mathcal{E}[Y]$。
- 对称性　$C_{XY}=C_{YX}$。
- 展缩和平移　$\mathrm{cov}[aX+b,\ cY+d]=acC_{XY}$，其中 $a,b,c,d \in \mathcal{R}$ 是常数。

这些性质的证明留作练习。

例 5-29 假设 X_1 和 X_2 是均值相同的相关随机变量，定义 $Y_1 \triangleq X_1+X_2$、$Y_2 \triangleq X_1-X_2$。由于 $\mu_{Y_2}=\mathcal{E}[X_1-X_2]=0$，$Y_1$ 和 Y_2 的协方差等于：

$$\begin{aligned} C_{Y_1Y_2} &= \mathcal{E}[(X_1+X_2)(X_1-X_2)] - \mu_{Y_1}\mu_{Y_2} \\ &= \mathcal{E}[X_1^2] + \mathcal{E}[X_2^2] = R_{Y_1Y_2} \end{aligned} \tag{5-217}$$

其中，$\mu_{Y_2}=\mathcal{E}[X_1-X_2]=0$。 ◀

需要注意的是当 $Y=X$ 时，$C_{XY}\to C_{XX}=\sigma_X^2$，$R_{XY}\to R_{XX}=\mathcal{E}[X^2]$。虽然符号 C_{XX} 和 R_{XX}

不是标准的随机变量符号，但它们将用于第 6 章随机过程。那里它们取决于过程中的时刻 $\{t_1,t_2\}$：$C_{XX}(t_1,t_2)\triangleq\mathcal{E}[[X(t_1)-\mu_X(t_1)][X(t_2)-\mu_X(t_2)]]$、$R_{XX}(t_1,t_2)\triangleq\mathcal{E}[X(t_1)X(t_2)]$，分别称为自协方差和自相关函数。

5.16 方差的函数观点

利用类似解释均值的函数方法，来说明如何利用方差度量概率密度函数的宽度。由于计算方差时已减去均值使得概率密度函数的中心移到 0，因此它是一个中心矩。这里假设 X 的均值为零，从而研究

$$\mathcal{E}[X^2]=\int_{-\infty}^{\infty}x^2 f_X(x)\mathrm{d}x \tag{5-218}$$

用 $f_X(x)$ 表示参数 $\{\mu=0,\ \sigma=1\}$ 的标准高斯分布。图 5-12a 画出了乘积 $x^2 f_X(x)$（虚线）。由于 x^2 和 $f_X(x)$ 非负，式(5-218)也必须非负。如果概率密度函数变宽（σ 变大），因为 x^2 是一个快速增长的非负函数，$x^2 f_X(x)$ 将有更大的面积。从图 5-12b 可以看出当 $\sigma^2=2$ 时，虚线下的面积增大。与用于计算均值的 $xf_X(x)$ 中的 x 不同，x^2 没有零交叉，它是关于原点的偶函数，因此 $f_X(x)$ 在 $x=0$ 两侧有相等的正加权。图 5-12 中画出了累积面积

$$B(x)=\int_{-\infty}^{x}v^2 f_X(v)\mathrm{d}v \tag{5-219}$$

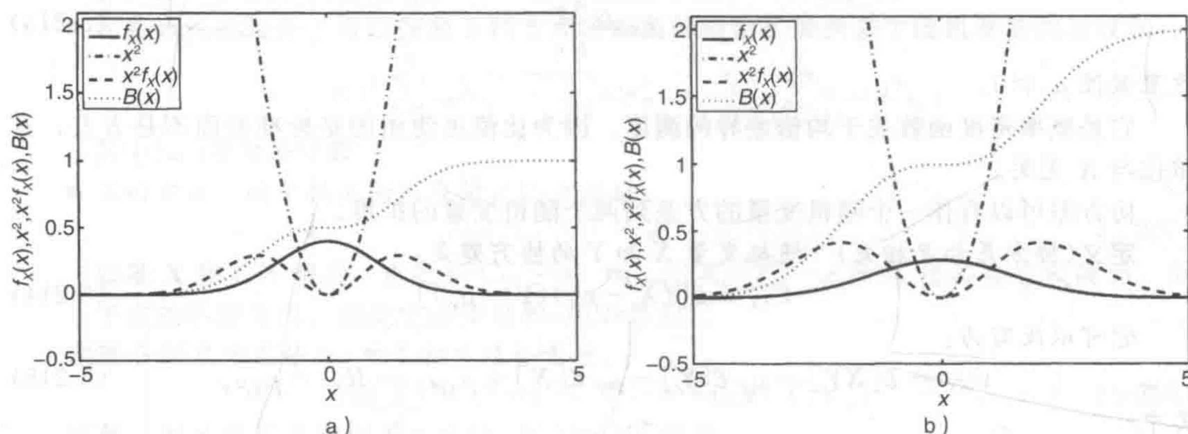


图 5-12 方差的函数观点画出了高斯概率密度函数的 $f_X(x)$，x^2，$x^2 f_X(x)$ 和 $B(x)$。a) $\mu=0$ 和 $\sigma=1$；b) $\mu=0$，$\sigma=\sqrt{2}$

x 是取值范围在 $f_X(x)$ 定义域的变量。当然，$\mathcal{E}[X^2]=\lim\limits_{x\to\infty}B(x)$。该函数类似累积分布函数，但是被积函数是加权了的概率密度函数 $x^2 f_X(x)$。

从该结果可以得出结论：均值的其他偶函数同样是概率密度函数宽度的测度。这方面的例子包括幂次是偶数的高阶中心矩，例如 $\mathcal{E}[(X-\mu_X)^4]$，$\mathcal{E}[(X-\mu_X)^6]$ 等，还有 $\mathcal{E}[|X-\mu_X|]$。当 $\mu_X=0$ 时，$\mathcal{E}[|X-\mu_X|]$ 变成

$$\mathcal{E}[|X|]=\int_{-\infty}^{\infty}|x|f_X(x)\mathrm{d}x=\int_{0}^{\infty}xf_X(x)\mathrm{d}x-\int_{-\infty}^{0}xf_X(x)\mathrm{d}x \tag{5-220}$$

改变第二个积分中的变量得到：

$$\mathcal{E}[|X|]=\int_{0}^{\infty}x[f_X(x)+f_X(-x)]\mathrm{d}x \tag{5-221}$$

由于 $x\geqslant0$、$f_X(x)$ 非负，如果 $f_X(x)$ 变宽，$x[f_X(x)+f_X(-x)]$ 下的面积总是为正且随之增加，虽然增加速度没有 $x^2 f_X(x)$ 快。

5.17 期望和指示函数

有时可以利用指示函数方便地将各种概率量写成期望的形式。对于概率密度函数为 $f_X(x)$ 和累积分布函数为 $F_X(x)$ 的随机变量 X：

$$P(a \leqslant X \leqslant b) = \int_{-\infty}^{\infty} I_{[a,b]}(x) f_X(x) \mathrm{d}x = \mathcal{E}[I_{[a,b]}(X)] \tag{5-222}$$

这种情况下，$I_{[a,b]}(X)$ 是随机变量 X 的特殊的非线性变换。类似地，可以将累积分布函数写成如下形式：

$$F_X(x) = P(X \leqslant x) = \mathcal{E}[I_{(-\infty,x]}(X)] \tag{5-223}$$

我们可能对指示函数本身的各种期望感兴趣。例如：

$$\begin{aligned}\operatorname{var}[I_{[a,b]}(X)] &= \mathcal{E}[(I_{[a,b]}(X) - \mathcal{E}[I_{[a,b]}(X)])^2] \\ &= \mathcal{E}[I_{[a,b]}^2(X)] - P^2(a \leqslant X \leqslant b)\end{aligned} \tag{5-224}$$

这里利用式(5-222)取代指示函数的均值。由于 $I_{[a,b]}(X)$ 的平方给出了同样的指示函数，最终得到：

$$\begin{aligned}\operatorname{var}[I_{[a,b]}(X)] &= P(a \leqslant X \leqslant b) - P^2(a \leqslant X \leqslant b) \\ &= P(a \leqslant X \leqslant b)[(1 - P(a \leqslant X \leqslant b)]\end{aligned} \tag{5-225}$$

与随机变量 X 的任意函数一样，计算方差时可以不计算映射的概率密度函数。上面结果的一个特殊情况是：

$$\operatorname{var}[I_{(-\infty,x]}(X)] = F_X(x)[1 - F_X(x)] \tag{5-226}$$

对于随机变量 X 和 Y，可以利用一系列操作计算协方差：

$$\begin{aligned}\operatorname{cov}[I_{[a,b]}(X) I_{[c,d]}(Y)] &= \mathcal{E}[I_{[a,b)}(X) I_{[c,d]}(Y)] - \mathcal{E}[I_{[a,b]}(X)]\mathcal{E}[I_{[c,d]}(Y)] \\ &= P(\{a \leqslant X \leqslant b) \cap \{c \leqslant Y \leqslant d)) - P(a \leqslant X \leqslant b) \\ &\quad P(c \leqslant Y \leqslant d)\end{aligned} \tag{5-227}$$

由于指示函数只有两个值 $\{0,1\}$，因此它的期望非负并且不超过 1，从式(5-222)中的 $0 \leqslant P(a \leqslant X \leqslant b) \leqslant 1$ 可以清楚地看出这一点。同样地，指示函数的方差有界：从式(5-225)得到：

$$0 \leqslant P(a \leqslant X \leqslant b)[1 - P(a \leqslant X \leqslant b)] \leqslant 1 - P(a \leqslant X \leqslant b) \leqslant 1 \tag{5-228}$$

式(5-227)的协方差同样有界。

例 5-30 设 X 服从参数 $\{\mu=0, \sigma=1\}$ 的标准高斯分布。那么

$$\mathcal{E}[I_{[0,\infty)}(X)] = 1/2 \tag{5-229}$$

以及

$$\operatorname{var}[I_{[0,\infty)}(X)] = P(0 \leqslant X < \infty)[1 - P(0 \leqslant X < \infty)] = 1/4 \tag{5-230}$$ ◀

5.18 相关系数

相关是两个或多个随机变量关系的一个重要测度。例如，如果 X 和 Y 是相关的，那么 X 的结果可以用于预测 Y 的结果。相关也意味着，如果 Y 是 X 的条件(反之亦然)，则 Y 给定时 X 的条件概率密度函数与边缘概率密度函数不同。在第 4 章例 4-5 中的二维高斯概率密度函数已经讨论过这一点。本节将讨论相关的归一化测度，它也可以用于产生给定相关度的随机变量。

定义(皮尔逊相关) 随机变量 X 和 Y 的**皮尔逊相关系数**为：

$$\rho_{XY} \triangleq \frac{C_{XY}}{\sigma_X \sigma_Y} \tag{5-231}$$

当提到相关系数时，一般指的是皮尔逊相关系数 ρ_{XY}(不是后面介绍的其他相关系数)。由于进行了归一化，ρ_{XY} 无量纲且有界。

定理 5-8 相关系数是有界的：

$$|\rho_{XY}| \leqslant 1 \tag{5-232}$$

证明：利用附录 F 中的柯西-施瓦兹不等式：

$$|\mathcal{E}[(X-\mu_X)(Y-\mu_Y)]| \leqslant \sqrt{\mathcal{E}[X-\mu_X]\,\mathcal{E}[Y-\mu_Y]} = \sigma_X\sigma_Y \tag{5-233}$$

等式两边同时除以 $\sigma_X\sigma_Y$，定理得证。

假设 U 和 V 是均值为 0、方差为 1 的不相关的随机变量。为了产生一组相关系数 ρ_{XY} 为特定值的随机变量$\{X, Y\}$，可以利用下列变换：

$$X = U,\ \ Y = \rho_{XY}U + \sqrt{1-\rho_{XY}^2}V \tag{5-234}$$

因此 $\sigma_X^2=1$，$\sigma_Y^2=\rho_{XY}^2+1-\rho_{XY}^2=1$，以及

$$\mathcal{E}[XY] = \rho_{XY}\,\mathcal{E}[U^2] + \sqrt{1-p_{XY}}\,\mathcal{E}[UV] = \rho_{XY} \tag{5-235}$$

从而 X 和 Y 是均值为 0、方差为 1、相关系数为 ρ_{XY} 的随机变量。从散点图中可以看出当 ρ_{XY} 改变时，相关随机变量是如何改变的。具体过程是先产生$\{u, v\}$，利用式(5-234)进行变换，并且绘制多个随机对。图 5-13 给出了多个 ρ_{XY} 的例子，其中$\{U, V\}$是独立的标准高斯随机变量。观察发现“星群”从圆形($\rho_{XY}=0$)变化到直线($\rho_{XY}=\pm1$)。图 5-14 给出了这些图形对应的 4 个概率密度函数，从中可以发现同一概率密度函数值对应的等高线与散点图的形状相同。它们从圆形($\rho_{XY}=0$)变化到椭圆($0<|\rho_{XY}|<1$)，然后到直线($|\rho_{XY}|=1$)。(因为不易画出概率密度函数为垂直平面的三维图，所以没画出直线情况。)

但是需注意相关系数不能描述两个随机变量所有的相关类型，例如图 5-13 中的线性相关。接下来的简单例子说明了这一点。

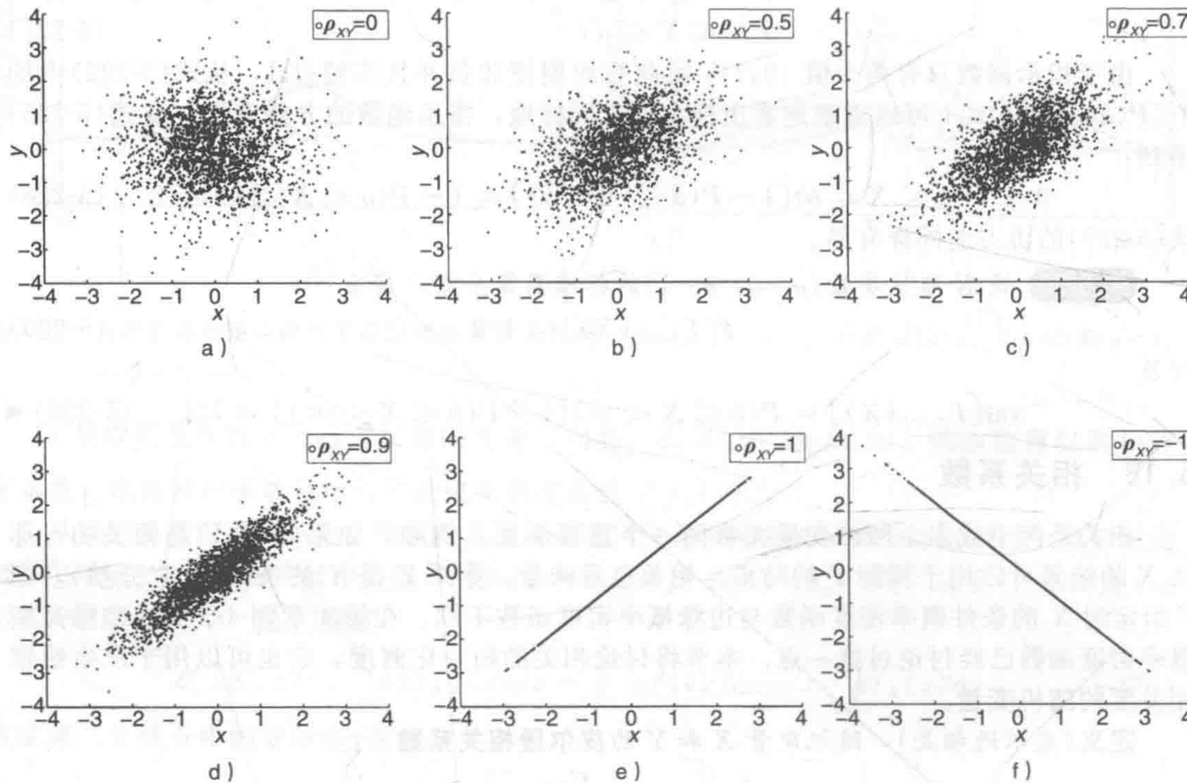


图 5-13 相关随机变量$\{X, Y\}$的不同相关系数对应的散点图。a) $\rho_{XY}=0$；b) $\rho_{XY}=0.5$；c) $\rho_{XY}=0.7$；d) $\rho_{XY}=0.9$；e) $\rho_{XY}=1$；f) $\rho_{XY}=-1$

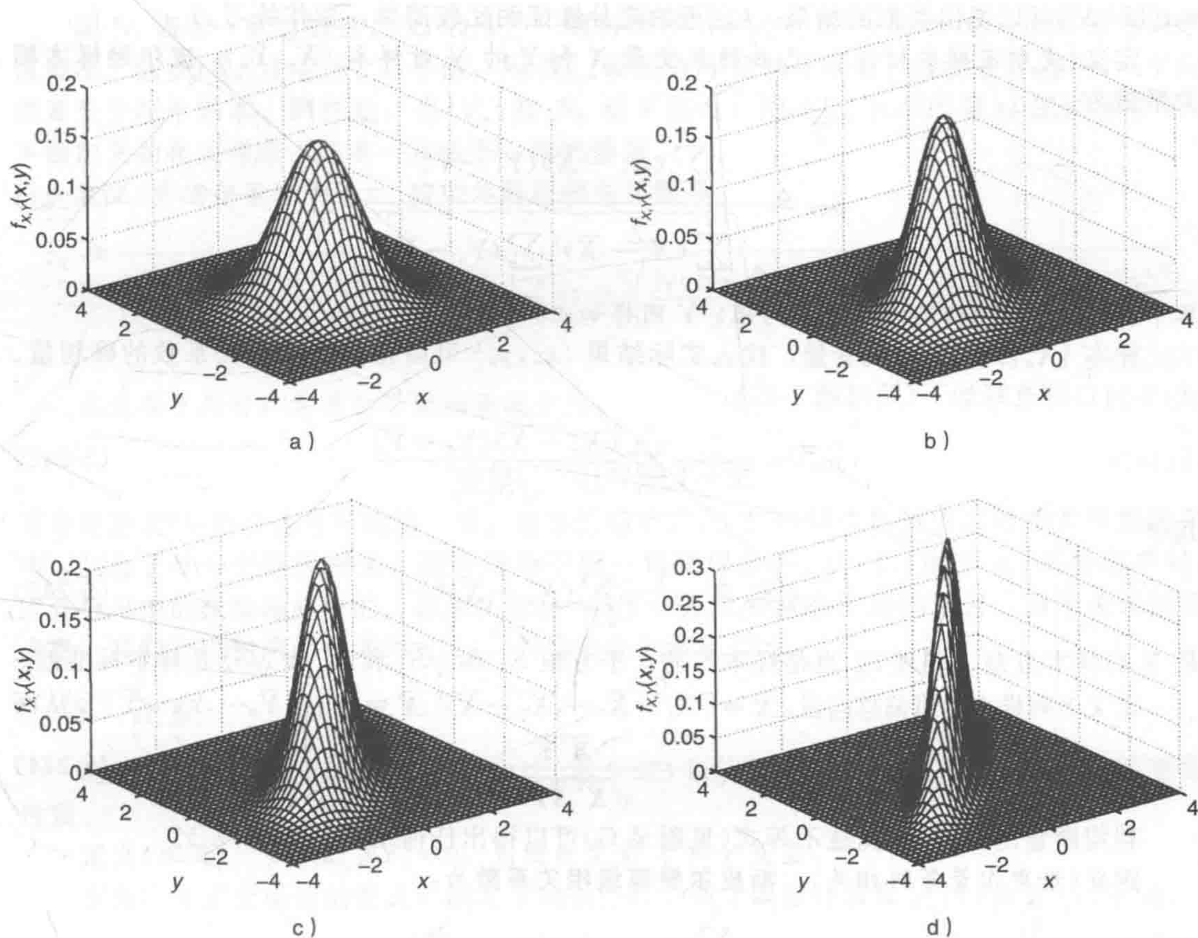


图 5-14 相关随机变量 $\{X,Y\}$ 的不同相关系数对应的二元高斯概率密度函数。a) $\rho_{XY}=0$；b) $\rho_{XY}=0.5$；c) $\rho_{XY}=0.7$；(d) $\rho_{XY}=0.9$

例 5-31 假设 X 是结果为$\{-1, 1\}$的对称伯努利随机变量，成功概率 $p=1/2$。变换 $Y=|X|$ 的概率密度函数 $f_Y(y)=\delta(x-1)$，从而

$$f_{X,Y}(x,y)=\begin{cases}1/2, & x=-1, y=1\\ 1/2, & x=1, y=1\\ 0, & \text{其他}\end{cases} \tag{5-236}$$

显然 X 和 Y 不独立，但是它们不相关。

$$\mathcal{E}[XY]=1/2\times(-1)\times 1+1/2\times 1\times 1=0 \tag{5-237}$$

因此 $\rho_{XY}=0$。 ◀

下面的定理描述了随机变量之间的线性关系。

定理 5-9 当且仅当 $Y=aX+b$ 时，$|\rho_{XY}|=1$。其中$\{a, b\}\in\mathcal{R}$是常数。

证明： 必要性：定义 $U\triangleq X/\sigma_X$、$V\triangleq Y/\sigma_Y$，并考虑差的方差：

$$\text{var}[U-Y]=\sigma_U^2+\sigma_V^2-2C_{UV}=2(1-\rho_{UV}) \tag{5-238}$$

则

$$\sigma_U^2=\mathcal{E}[(U-\mu_U)^2]=\mathcal{E}[(X-\mu_X)^2]/\sigma_X^2=1 \tag{5-239}$$

同理 $\sigma_V^2=1$。当 $\rho_{UV}=\rho_{XY}=1$ 时，$\text{var}[U-V]=0$，这也意味着 $U-V$ 只有一个值 $c\in\mathcal{R}$；概率密度函数 $f_{U-V}(z)=\delta(z-c)$。因此，$U-V$ 等于 c 的概率为 1，并且

$$Y=\sigma_Y(U-c)=(\sigma_Y/\sigma_X)X-c\sigma_Y \tag{5-240}$$

当然它是斜率 $a=\sigma_Y/\sigma_X$、截距 $b=-c\sigma_Y$ 的线性(仿射)函数。当 $\rho_{UV}=\rho_{XY}=-1$ 时，利用

var[$U+V$]可以得出类似的结果。(定理的充分性证明比较简单，留作练习。)

定义(皮尔逊样本相关) 已知随机变量 X 和 Y 的 N 对样本 $\{X_n,Y_n\}$，**皮尔逊样本相关系数**为：

$$\bar{\rho}_{XY} \triangleq \frac{\sum_{n=1}^{N}(X_n-\overline{X})(Y_n-\overline{Y})}{\sqrt{\sum_{n=1}^{N}(X_n-\overline{X})^2\sum_{n=1}^{N}(Y_n-\overline{Y})^2}} \tag{5-241}$$

其中 $\overline{X}$ 是式(4-228)定义的样本均值，$\overline{Y}$ 同样如此。

样本 $\{X_n,Y_n\}$ 是随机变量，代入实际结果 $\{x_n,y_n\}$ 可以得到样本相关系数的确切值。式(5-241)可改写如下(见习题 5-33)：

$$\bar{\rho}_{XY} = \frac{1}{N-1}\sum_{n=1}^{N}\frac{(X_n-\overline{X})(Y_n-\overline{Y})}{\sqrt{S_X^2S_Y^2}} \tag{5-242}$$

其中

$$S_X^2 \triangleq \frac{1}{N-1}\sum_{n=1}^{N}(X_n-\overline{X})^2 \tag{5-243}$$

是 X 的样本方差，同理 S_Y^2 也是样本方差。平方根 $S_X \triangleq \sqrt{S_X^2}$ 和 $S_Y \triangleq \sqrt{S_Y^2}$ 是样本标准差。

定义下列样本均值偏移向量：$\widetilde{\boldsymbol{X}}=[X_1-\overline{X},\cdots,X_N-\overline{X}]^{\mathrm{T}}$、$\widetilde{\boldsymbol{Y}}=[Y_1-\overline{Y},\cdots,Y_N-\overline{Y}]^{\mathrm{T}}$，从而

$$\bar{\rho}_{XY} = \frac{\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{Y}}}{\sqrt{\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{Y}}^{\mathrm{T}}\widetilde{\boldsymbol{Y}}}} \tag{5-244}$$

利用向量的柯西-施瓦兹不等式(见附录 G)可以得出任何结果的 $|\bar{\rho}_{XY}|\leqslant 1$。

定义(斯皮尔曼等级相关) **斯皮尔曼等级相关系数**为：

$$\tau_S \triangleq \frac{\sum_{n=1}^{N}(r_{X,n}-\bar{r}_X)(r_{Y,n}-\bar{r}_Y)}{\sqrt{\sum_{n=1}^{N}(r_{Y,n}-\bar{r}_X)^2\sum_{n=1}^{N}(r_{Y,n}-\bar{r}_Y)^2}} \tag{5-245}$$

其中 $\{r_{X,n}\}$ 和 $\{r_{Y,n}\}$ 分别是样本 $\{X_n\}$ 和 $\{Y_n\}$ 的等级。等级对应的均值分别为 $\bar{r}_X$ 和 $\bar{r}_Y$。

特别要注意，样本 $\{X_n\}$ 和 $\{Y_n\}$ 的实际值只用于计算等级，而不会用于计算 τ_S。尽管式(5-245)的向量比包含等级，同样可以借助柯西-施瓦兹不等式得出斯皮尔曼等级相关系数也是有界的，即 $|\tau_S|\leqslant 1$。

在介绍随机变量样本的最后一个相关测度前，先根据样本等级给出以下定义。

定义(一致和不一致) 设 $\{X_n,\ Y_n\}$ 是随机变量 X 和 Y 的样本，如果

$$\{X_n<X_m,Y_n<Y_m\} \text{ 或 } \{X_n>X_m,Y_n>Y_m\} \tag{5-246}$$

$\{X_n,Y_n\}$，$\{X_m,Y_m\}$ 是一致的。

如果

$$\{X_n<X_m,Y_n>Y_m\} \text{ 或 } \{X_n>X_m,Y_n<Y_m\} \tag{5-247}$$

它们是不一致的。

如果

$$\{X_n=X_m\} \text{ 或 } \{Y_n=Y_m\} \tag{5-248}$$

它们既不是一致的，也不是不一致的。

包含等号的最后一个条件称为平局。排除一致或不一致时，只需要其中的一个等式。N 个样本中，组合对(N 选 2)的数目为：

$$\binom{N}{2}=\frac{N!}{2!(N-2)!}=N(N-1)/2 \tag{5-249}$$

用 N_c 表示一致对数目，N_d 表示不一致对数目，定义差 $\Delta_{cd} \triangleq N_c - N_d$。同样要考虑平值概率。假设 $\{X_n\}$ 中有 N_X 组平值，用 $N_{X,m}$ 表示第 m 组的样本数。显然，$N_{X,m} \geqslant 2$：平局需要至少两个样本。同样地，当 $\{Y_n\}$ 有 N_Y 组平值时，用 $N_{Y,m}$ 表示的第 m 组的样本数。下面定义的相关考虑了样本一致或不一致的情况。

定义（肯德尔等级相关） **肯德尔等级相关系数**为：

$$\tau_K \triangleq \frac{\Delta_{cd}}{\left(\left[\frac{N(N-1)}{2} - \sum_{m=1}^{N_X} N_{X,m}(N_{X,m}-1)/2\right]\left[\frac{N(N-1)}{2} - \sum_{m=1}^{N_Y} N_{Y,m}(N_{Y,m}-1)/2\right]\right)^{1/2}} \tag{5-250}$$

当没有平局时，肯德尔等级相关减少为：

$$\tau_K = \frac{\Delta_{cd}}{N(N-1)/2} \tag{5-251}$$

其分母是式(5-250)的分母的第一项。这里已经将式(5-250)写成类似皮尔逊相关系数的形式，其分子是一个等级相关，而分母用于归一化来保证 $|\tau_K| \leqslant 1$。由于 Δ_{cd} 不考虑平局，因此分母中的减法是必要的。必须从组合 $N(N-1)/2$ 中减去平局的数目。对于连续随机变量，我们可以使用简化式(5-251)，因为平局的概率为 0。

5.19 正交

两个随机变量的正交性在参数估计（第 9 章）和信号估计（第 11 章）时是非常重要的性质。

定义（正交） 如果 $\mathcal{E}[XY]=0$，则随机变量 X 和 Y **正交**。

根据广义正交函数的形式：满足下列条件时，两个确定性函数 $g_1(x)$ 和 $g_2(x)$ 正交：

$$\int_{-\infty}^{\infty} g_1(x) g_2(x) w(x) \mathrm{d}x = 0 \tag{5-252}$$

其中 $w(x)$ 是一个正的加权函数。这里将加权函数写成 X 和 Y 的联合概率密度函数，积分变量是两个随机变量：

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y) \mathrm{d}x\mathrm{d}y = 0 \tag{5-253}$$

下面的定理说明了两个随机变量独立、不相关、正交的联系。

定理 5-10 对于随机变量 X 和 Y：

- 独立意味着不相关：

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \Rightarrow \mathcal{E}(XY] = \mathcal{E}[X]\,\mathcal{E}[Y] \tag{5-254}$$

- 当至少一个随机变量的均值为 0 时，不相关和正交等价：

$$\mathcal{E}[XY] = \mathcal{E}[X]\,\mathcal{E}[Y], \quad \mu_X = 0 \text{ 或 } \mu_Y = 0 \Rightarrow \mathcal{E}[XY] = 0 \tag{5-255}$$

证明： 前面的叙述已经证明了该定理。

虽然零均值随机变量的不相关和正交等价，但它们是不同的特性。在许多应用中，由于问题的性质或零均值随机变量的变换的应用，使得随机变量均值为 0，因此经常将这两个性质混淆。由式(5-253)发现，正交性取决于概率密度函数与 xy 乘积关系的结构：函数 $xyf_{X,Y}(x, y)$ 在区间内的正、负面积是否相等。因此，当 X 和 Y 都是正或都是负时，它们不可能正交。

另一方面，由于相关是两个随机变量线性关系测度，不相关只意味着不存在线性关系。图 5-13 中，当相关系数 ρ 接近 ± 1 时，这种线性关系是显而易见的。通常，相关随机变量意味着不独立。然而，从下面的例子可以发现有些分布是独立的。

例 5-32 设 X 和 Y 是联合高斯随机变量，二元概率密度函数为：

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp(-((x-\mu_X)^2/\sigma_X^2 - 2\rho(x-\mu_X)(y-\mu_Y)/\sigma_X\sigma_Y+(y-\mu_Y)^2/\sigma_Y^2)/2(1-\rho_{XY}^2)) \quad (5\text{-}256)$$

如果 X 和 Y 不相关($\rho_{XY}=0$)，则 $f_{X,Y}(x,y)$可表示为两个边缘概率密度函数的乘积：

$$\begin{aligned}f_{X,Y}(x,y)&=\frac{1}{\sqrt{2\pi}\sigma_X}\exp(-(x-\mu_X)^2/2\sigma_X^2)\frac{1}{\sqrt{2\pi}\sigma_Y}\exp(-(y-\mu_Y)^2/2\sigma_Y^2)\\&=f_X(x)f_Y(y)\end{aligned} \quad (5\text{-}257)$$

虽然不相关高斯随机变量通常是独立的，但也有例外。例外通常是人为构造的变量，如例 5-33。◀

例 5-33 设 X 是参数$\{\mu=0，\sigma\}$的高斯随机变量。假设 $Y=ZX$，其中 Z 是对称的结果为$\{-1，1\}$的伯努利随机变量，$p=P(Z=1)=1-q$，且 Z 与 X 独立。Y 的累积分布函数为：

$$\begin{aligned}F_Y(y)&=P(Y\leqslant y|Z=1)p+P(Y\leqslant y|Z=-1)q\\&=P(X\leqslant y|Z=1)p+P(X\geqslant -y|Z=-1)q\end{aligned} \quad (5\text{-}258)$$

由于 X 和 Z 独立，两个条件概率可写成：

$$F_Y(y)=P(X\leqslant y)p+P(X\geqslant -y)q=P(X\leqslant y)=F_X(y) \quad (5\text{-}259)$$

这里利用了 $f_X(x)$是偶函数。因此，Y 和 X 的分布相同。显然，X 和 Y 不独立：当 $X=x$ 时，Y 必须是$+x$ 或$-x$。另一方面，当 $p=q=1/2$ 时，它们是不相关的：

$$C_{XY}=\mathcal{E}[XY]=\mathcal{E}[X^2Z]=\mathcal{E}[X^2]\mathcal{E}[Z]=\sigma^2(p-q)=0 \quad (5\text{-}260)$$

虽然 X 和 Y 各自服从高斯分布，但它们不服从联合高斯分布。这也是它们不相关但不独立的原因。◀

例 5-34 接下来，我们希望得到例 5-33 中随机变量 X 和 Y 的联合概率密度函数。首先假设 $p=1/2$，联合累积分布函数为：

$$\begin{aligned}F_{X,Y}(x,y)&=P(X\leqslant x,XZ\leqslant y|Z=1)P(Z=1)+P(X\leqslant x,XZ\leqslant y|Z=-1)P(Z=-1)\\&=(1/2)P(X\leqslant x,X\leqslant y)+(1/2)P(X\leqslant x,X\geqslant -y)\\&=(1/2)F_X(\min(x,y))+(1/2)[F_X(x)-F_X(-y)]u(x+y)\end{aligned} \quad (5\text{-}261)$$

其中阶跃函数使得只有当 $x\geqslant -y$ 时，第二项非零。第二项也可以写成：

$$F_X(x)-F_X(-y)=\int_{-y}^{x}f_X(x)\mathrm{d}x \quad (5\text{-}262)$$

其中，$f_X(x)$是标准高斯概率密度函数。对于独立随机变量，该累积分布函数与$F_X(x)F_Y(y)$的乘积有明显的不同。图 5-15 分别画出了第一项、第二项以及整体的累积分布函数。观察发现，$\min(x，y)$导致图 5-15a 中 $x=y$ 处出现“脊”，$u(x+y)$导致图 5-15b 中的 $x=-y$ 一侧突然出现平坦区域。当 $x<-y$ 时，函数 $F_X(x)-F_X(-y)$将为负值，但单位阶跃函数对它进行了截断，所以 $F_X(x)-F_X(-y)$总是非负，从而产生了图中零处的“基底”。式(5-261)中的第二项可重写为：

$$F_X(x)-F_X(-y)=F_X(x)+F_X(y)-1 \quad (5\text{-}263)$$

表明它关于 x 和 y 对称。利用下式推导出式(5-261)的概率密度函数：

$$\min(x,y)=(1/2)(x+y-|x-y|) \quad (5\text{-}264)$$

从而

$$\frac{\partial}{\partial x}\min(x,y)=(1/2)[1-\mathrm{sgn}(x-y)],\quad \frac{\partial^2}{\partial x\partial y}\min(x,y)=\delta(x-y) \quad (5\text{-}265)$$

其中 $\mathrm{dsgn}(y)/\mathrm{d}y=2\delta(y)$。受累积分布函数中第一项的影响，整体的概率密度函数包含 $x=y$线处的冲激函数(见习题 5-37)。

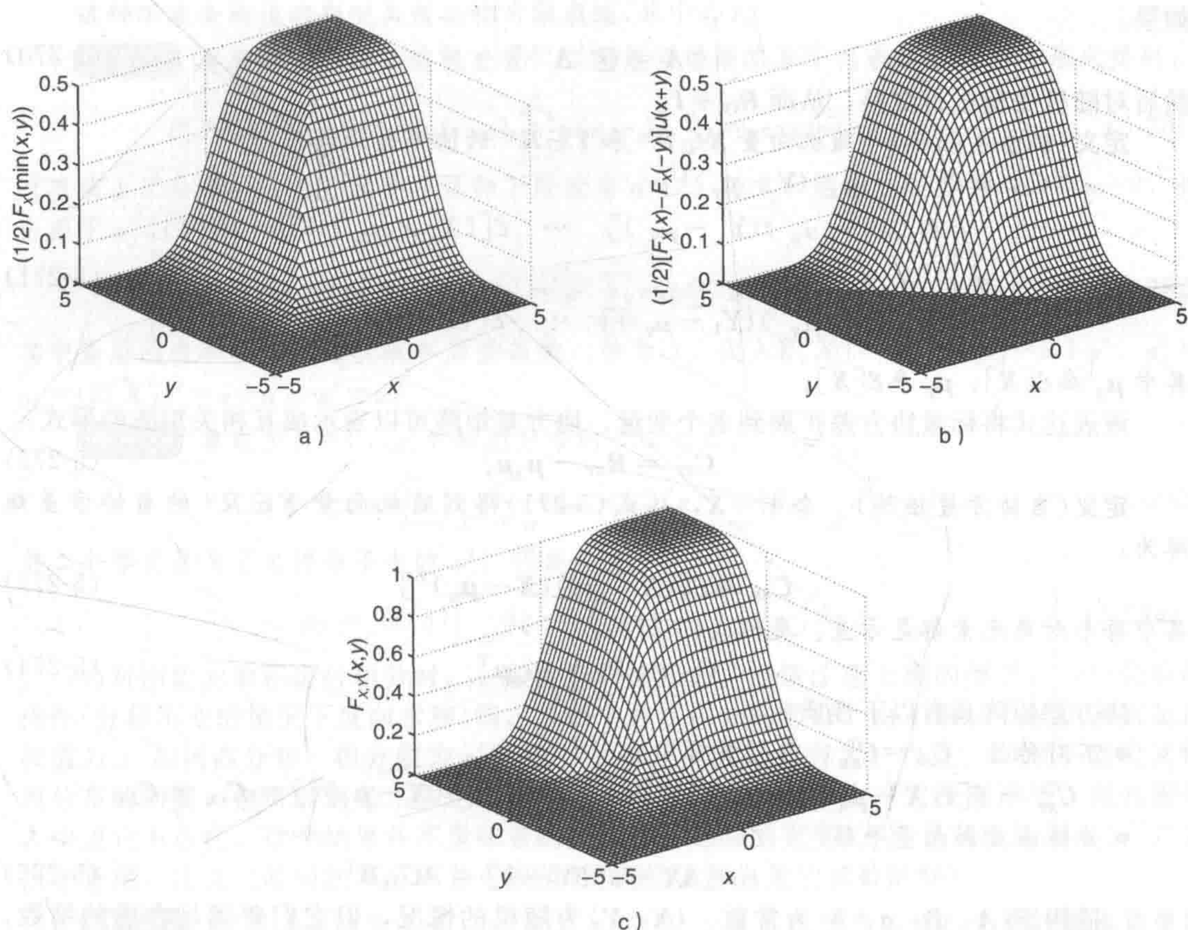


图 5-15 例 5-34 不相关随机变量的累积分布函数。a) 第一项：$(1/2)F_X(\min(x,y))$；b) 第二项：$(1/2)[F_X(x)-F_X(-y)]u(x+y)$；c) a 和 b 的总和 ◀

5.20 相关和协方差矩阵

考虑包含 N 个随机变量$\{X_n\}$的随机向量 $\boldsymbol{X}=[X_1,\cdots,X_N]^{\mathrm{T}}$。

定义(自相关矩阵) 随机向量 $\boldsymbol{X}\in\mathcal{R}^N$ 的自相关矩阵为：

$$\boldsymbol{R_{XX}}\triangleq\mathcal{E}[\boldsymbol{XX}^{\mathrm{T}}]=\begin{pmatrix}\mathcal{E}[X_1^2] & \mathcal{E}[X_1X_2] & \cdots & \mathcal{E}[X_1X_N]\\ \mathcal{E}[X_2X_1] & \mathcal{E}[X_2^2] & & \\ \vdots & & \ddots & \vdots\\ \mathcal{E}[X_NX_1] & \cdots & & \mathcal{E}[X_N^2]\end{pmatrix} \tag{5-266}$$

第(m, n)项为$\mathcal{E}[X_mX_n]$，其中 m 是行数、n 为列数。

设$\boldsymbol{Y}=[Y_1,\cdots,Y_M]^{\mathrm{T}}$ 包含 M 个随机变量$\{Y_M\}$。

定义(互相关矩阵) 随机向量 $\boldsymbol{X}\in\mathcal{R}^N$ 和 $\boldsymbol{Y}\in\mathcal{R}^N$ 的互相关矩阵为：

$$\boldsymbol{R_{XY}}\triangleq\mathcal{E}[\boldsymbol{XY}^{\mathrm{T}}] \tag{5-267}$$

假设随机向量的随机变量在一定程度上相关。通过乘以变换矩阵，可以将随机向量的元素变成一组新的互不相关的随机变量。考虑变换 $\boldsymbol{Y}=\boldsymbol{AX}$，其中 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$，因此

$$\boldsymbol{R_{YY}}=\mathcal{E}[\boldsymbol{AXX}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}]=\boldsymbol{AR_{XX}A}^{\mathrm{T}} \tag{5-268}$$

如果 $\boldsymbol{R_{YY}}$是对角矩阵，则 $\boldsymbol{Y}$ 的元素不相关。这是通过特征分解(见附录 G)得到的：

$$\boldsymbol{R_{XX}}=\boldsymbol{Q\Lambda Q}^{\mathrm{T}} \tag{5-269}$$

其中对角矩阵 $\boldsymbol{\Lambda}$ 中的元素为 $\boldsymbol{R_{XX}}$ 的特征值。$\boldsymbol{Q}$ 的列是相应的特征向量，因此 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}$，其中 $\boldsymbol{I}\in R^{N\times N}$ 是单位矩阵，因此 $\boldsymbol{Q}^{-1}=\boldsymbol{Q}^{\mathrm{T}}$。令 $\boldsymbol{A}=\boldsymbol{Q}^{-1}$，得到需要的对角结果：$\boldsymbol{R_{XY}}=\boldsymbol{\Lambda}$。

如果：

$$\boldsymbol{A} = \boldsymbol{Q}^{-1}\boldsymbol{\Lambda}^{-1/2} \tag{5-270}$$

然后对随机变量进行变换，从而 $\boldsymbol{R}_{YY}=\boldsymbol{I}$。

定义(协方差矩阵)　随机向量 $\boldsymbol{X}\in\mathcal{R}^N$ 和 $\boldsymbol{Y}\in\mathcal{R}^M$ 的**协方差矩阵**为：

$$\begin{aligned}\boldsymbol{C}_{XY} &\triangleq \mathcal{E}[(\boldsymbol{X}-\boldsymbol{\mu}_X)(\boldsymbol{Y}-\boldsymbol{\mu}_Y)^{\mathrm{T}}] \\ &= \begin{bmatrix} \mathcal{E}[(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})] & \cdots & \mathcal{E}[(X_1-\mu_{X_1})(Y_M-\mu_{Y_M})] \\ \vdots & \ddots & \vdots \\ \mathcal{E}[(X_N-\mu_{X_N})(Y_1-\mu_{Y_1})] & \cdots & \mathcal{E}[(X_N-\mu_{X_N})(Y-\mu_{Y_M})] \end{bmatrix}\end{aligned} \tag{5-271}$$

其中 $\boldsymbol{\mu}_X\triangleq\mathcal{E}[\boldsymbol{X}]$，$\boldsymbol{\mu}_Y\triangleq\mathcal{E}[\boldsymbol{X}]$。

该表达式将标量协方差扩展到多个变量。协方差矩阵可以表示成互相关矩阵的形式：

$$\boldsymbol{C}_{XY} = \boldsymbol{R}_{XY} - \boldsymbol{\mu}_X\boldsymbol{\mu}_Y^{\mathrm{T}} \tag{5-272}$$

定义(自协方差矩阵)　令 $\boldsymbol{Y}=\boldsymbol{X}$，从式(5-271)得到随机向量 $\boldsymbol{X}\in\mathcal{R}^N$ 的自协方差矩阵为：

$$\boldsymbol{C}_{XX} \triangleq \mathcal{E}[(\boldsymbol{X}-\boldsymbol{\mu}_X)(\boldsymbol{X}-\boldsymbol{\mu}_X)^{\mathrm{T}}] \tag{5-273}$$

其中每个对角元素都是方差。类似式(5-272)：

$$\boldsymbol{C}_{XX} = \boldsymbol{R}_{XX} - \boldsymbol{\mu}_X\boldsymbol{\mu}_X^{\mathrm{T}} \tag{5-274}$$

协方差矩阵具有以下性质：

- 不对称性　$\boldsymbol{C}_{XY}=\boldsymbol{C}_{YX}^{\mathrm{T}}$，由以下可知：

$$\boldsymbol{C}_{XY}^{\mathrm{T}} = \mathcal{E}[((\boldsymbol{X}-\boldsymbol{\mu}_X)(\boldsymbol{Y}-\boldsymbol{\mu}_Y)^{\mathrm{T}})^{\mathrm{T}}] = \mathcal{E}[(\boldsymbol{Y}-\boldsymbol{\mu}_Y)(\boldsymbol{X}-\boldsymbol{\mu}_X)^{\mathrm{T}}] = \boldsymbol{C}_{YX} \neq \boldsymbol{C}_{XY}$$

- 矩阵缩放和向量平移

$$\operatorname{cov}[\boldsymbol{AX}+\boldsymbol{a},\boldsymbol{BY}+\boldsymbol{b}] = \boldsymbol{AC}_{XY}\boldsymbol{B}^{\mathrm{T}} \tag{5-275}$$

适用于$\{\boldsymbol{A},\boldsymbol{B},\boldsymbol{a},\boldsymbol{b}\}$为常数、$\{\boldsymbol{X},\boldsymbol{Y}\}$为随机的情况，但它们要满足合适的维数。由于均值变为$\mathcal{E}[\boldsymbol{AX}+\boldsymbol{a}]=\boldsymbol{A\mu}_X+\boldsymbol{a}$、$\mathcal{E}[\boldsymbol{BY}+\boldsymbol{b}]=\boldsymbol{B\mu}_Y+\boldsymbol{b}$，从 $\boldsymbol{AX}+\boldsymbol{a}$ 和 $\boldsymbol{BY}+\boldsymbol{b}$ 中分别减去它们，得到：

$$\mathcal{E}[(\boldsymbol{AX}-\boldsymbol{A\mu}_X)(\boldsymbol{BY}-\boldsymbol{B\mu}_Y)^{\mathrm{T}}] = \boldsymbol{A}\,\mathcal{E}[(\boldsymbol{X}-\boldsymbol{\mu}_X)(\boldsymbol{Y}-\boldsymbol{\mu}_Y)^{\mathrm{T}}]\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{AC}_{XY}\boldsymbol{B}^{\mathrm{T}} \tag{5-276}$$

- 随机向量相加

$$\operatorname{cov}[\boldsymbol{X}+\boldsymbol{Z},\boldsymbol{Y}] = \boldsymbol{C}_{XY}+\boldsymbol{C}_{ZY} \tag{5-277}$$

对于满足合适维数的随机向量$\{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}$，由以下可知：

$$\begin{aligned}\mathcal{E}[(\boldsymbol{X}+\boldsymbol{Z}-(\boldsymbol{\mu}_X+\boldsymbol{\mu}_Z))(\boldsymbol{Y}-\boldsymbol{\mu}_Y)^{\mathrm{T}}] &= \mathcal{E}[(\boldsymbol{X}-\boldsymbol{\mu}_X)(\boldsymbol{Y}-\boldsymbol{\mu}_Y)^{\mathrm{T}}]+\mathcal{E}[(\boldsymbol{Z}-\boldsymbol{\mu}_Z)(\boldsymbol{Y}-\boldsymbol{\mu}_Y)^{\mathrm{T}}] \\ &= \boldsymbol{C}_{XY}+\boldsymbol{C}_{ZY}\end{aligned} \tag{5-278}$$

这些性质也适用于互相关矩阵 $\boldsymbol{R}_{XY}$。自协方差矩阵具有以下特有性质：

- 对称性　$\boldsymbol{C}_{XX}=\boldsymbol{C}_{XX}^{\mathrm{T}}$。
- 半正定性(PSD)　对于所有 $\boldsymbol{v}\neq 0$ 的向量：

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}_{XX}\boldsymbol{v} = \mathcal{E}[\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{\mu}_X)(\boldsymbol{X}-\boldsymbol{\mu}_X)^{\mathrm{T}}\boldsymbol{v}] = \mathcal{E}[(\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{\mu}_X))^2] \geqslant 0 \tag{5-279}$$

这意味着 $\boldsymbol{C}_{XX}$的所有特征值都非负。如果矩阵非奇异(有逆矩阵 $\boldsymbol{C}_{XX}^{-1}$)，则特征值为正值且 $\boldsymbol{C}_{XX}$正定(PD)(对于所有 $\boldsymbol{v}\neq 0$ 向量，式(5-279)取严格不等号“$>$”)。

这些性质也适用于自相关矩阵 $\boldsymbol{R}_{XX}$。

5.21　高阶矩和累积量

本节将均值和方差(一阶矩和二阶矩)的定义扩展到高阶矩。

定义(矩)　随机变量 X 的 n 阶矩为：

$$\mu_n \triangleq \mathcal{E}[X^n] \tag{5-280}$$

这种不减去均值的期望类型也称为**原点矩**(非中心)。

例 5-35 参数为 α 的泊松随机变量,二阶矩可借助推导下列表达式的闭合形式得到:

$$\mathcal{E}[X(X-1)] = \exp(-\alpha)\sum_{n=0}^{\infty} n(n-1)\frac{\alpha^n}{n!} = \exp(-\alpha)\sum_{n=2}^{\infty}\frac{\alpha^n}{(n-2)!} \tag{5-281}$$

这里消去了分母中的部分项并将求和下限改为 $n=2$。将求和中的变量替换为 $m=n-2$,提出因子 α^2,得到:

$$\mathcal{E}[X(X-1)] = \alpha^2\sum_{m=0}^{\infty}\exp(-\alpha)\frac{\alpha^m}{m!} = \alpha^2 \tag{5-282}$$

其中最后的求和项是有效的概率质量函数,值为1。代入 $\mathcal{E}[X]=\alpha$ 得到 $\mu_2=\alpha+\alpha^2$、$\sigma_X^2=\mu_2-(\mathcal{E}[X])^2=\alpha+\alpha^2-\alpha^2=\alpha$。 ◀

例 5-36 参数为 $\{c=0, \alpha\}$ 的柯西分布的二阶矩为:

$$\mu_2 = (\alpha/\pi)\int_{-\infty}^{\infty}\frac{x^2}{x^2+\alpha^2}\mathrm{d}x = (\alpha/\pi)\int_{-\infty}^{\infty}\left(1-\frac{\alpha^2}{x^2+\alpha^2}\right)\mathrm{d}x \tag{5-283}$$

第二个等式是为了去掉分子中的 x^2。因此:

$$\mu_2 = (\alpha/\pi)\int_{-\infty}^{\infty}\mathrm{d}x - (\alpha^3/\pi)\int_{-\infty}^{\infty}\frac{1}{x^2+\alpha^2}\mathrm{d}x = \infty \tag{5-284}$$ ◀

当对函数的乘积进行积分时,一般进行分部积分。但请注意上面的例子,一个简单的操作(分母不变的情况下反向求和)得出了两个容易积分的部分。由于第二个被积函数是加权值为 α^2 的柯西分布,积分值为 α^2。第一个积分肯定是无穷大。因此,即使不能定义柯西分布的均值,也可以定义二阶矩(非中心),但它为无穷大。由于它没有用处,因此附录A中说它不存在。这个结果并不意味着柯西分布有无限的宽度(方差是宽度的测度),只是由于重尾,定义二阶矩的积分不是有限值(前面给出的方差的函数解释)。

例 5-37 设 $\{X_n\}$ 是均值为 μ_X、方差为 σ_X^2 的 N 个独立同分布的随机变量,样本均值的二阶矩为:

$$\mathcal{E}[\overline{X}^2] = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\mathcal{E}[X_nX_m] = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1,m\neq n}^{N}\mathcal{E}[X_n]\mathcal{E}[X_m] + \frac{1}{N^2}\sum_{n=1}^{N}\mathcal{E}[X_n^2] \tag{5-285}$$

其中将 N^2 个求和分为两部分:一部分 $m\neq n$(N^2-N 项),另一部分 $m=n$(N 项)。第一个求和利用了独立条件,将期望写成乘积的形式。因此:

$$\mathcal{E}[\overline{X}^2] = (1/N^2)(N^2-N)\mu_X^2 + (1/N^2)N(\sigma_X^2+\mu_X^2) = \mu_X^2+\sigma_X^2/N \tag{5-286}$$

样本均值的期望是:

$$\mathcal{E}[\overline{X}] = \frac{1}{N}\sum_{n=1}^{N}\mathcal{E}[X_n] = \mu_X \tag{5-287}$$

这表明样本均值为 μ_X 的无偏估计。它的方差是:

$$\sigma_{\overline{X}}^2 = \mathcal{E}[\overline{X}^2] - \mathcal{E}^2[\overline{X}] = \sigma_X^2/N \tag{5-288}$$

这表明当样本数 N 增大时,$\overline{X}$ 的估计性能将提高(降低了方差)。后面的图9-8给出了一个高斯分布 $\{X_n\}$ 的结果的例子。 ◀

定义(中心矩) 均值为 μ_X 的随机变量 X 的 n 阶**中心矩**为:

$$\mu_{c,n} \triangleq \mathcal{E}[(X-\mu_X)^n] \tag{5-289}$$

方差是二阶中心矩:$\sigma_X^2 \equiv \mu_{c,2}$。概率密度函数关于均值对称的随机变量的奇数阶中心矩为0。根据定义很容易得出该结论:

$$\mu_{c,n} = \int_{-\infty}^{\infty}(x-\mu_X)^n f_X(x)\mathrm{d}x \tag{5-290}$$

观察发现,当 n 为偶数时,$(x-\mu_X)^n$ 是偶函数,当 n 为奇数时它是奇函数。由于偶函数(这种条件下的 $f_X(x)$)和奇函数的乘积为奇函数,当 n 为奇数时,乘积的面积肯定是

0(除非它不存在，如柯西分布)。

定理 5-11 参数为$\{\mu, \sigma\}$的随机变量X的n阶中心矩为：

$$\mu_{c,n}=\begin{cases}\sigma^n n!/2^{n/2}(n/2)!, & n\text{为偶数}\\ 0, & n\text{为奇数}\end{cases} \tag{5-291}$$

证明：(Papoulis, 1965) n为奇数的情况前面已经借助对称条件完成了证明。设$Y\triangleq X-\mu$，重写高斯概率密度函数面积的表达式如下(见附录D概率密度函数的积分)：

$$\int_{-\infty}^{\infty}\exp(-y^2/2\sigma^2)\mathrm{d}y=\sqrt{2\pi\sigma^2} \tag{5-292}$$

代入$a\triangleq 1/2\sigma^2$：

$$\int_{-\infty}^{\infty}\exp(-ay^2)\mathrm{d}y=\sqrt{\pi/a} \tag{5-293}$$

对等式两边反复关于a进行微分，当$n=2$，4，6时，结果为：

$$\int_{-\infty}^{\infty}(-y^2)\exp(-ay^2)\mathrm{d}y=a^{-3/2}(-1/2)\sqrt{\pi} \tag{5-294}$$

$$\int_{-\infty}^{\infty}y^4\exp(-ay^2)\mathrm{d}y=a^{-5/2}(1/2)(3/2)\sqrt{\pi} \tag{5-295}$$

$$\int_{-\infty}^{\infty}(-y^6)\exp(-ay^2)\mathrm{d}y=a^{-7/2}(1/2)(3/2)(-5/2)\sqrt{\pi} \tag{5-296}$$

由于两边的负号总是被抵消，根据n增大时的结果，得到通用表达式：

$$\int_{-\infty}^{\infty}y^{2n}\exp(-ay^2)\mathrm{d}y=a^{-(2n+1)/2}(1/2)^n\sqrt{\pi}\prod_{m=1}^{n}(2m-1) \tag{5-297}$$

该乘积仅包含奇正整数的阶乘，它被称为双阶乘，可改写为：

$$\prod_{m=1}^{n}(2m-1)=(2n)!/2^n n! \tag{5-298}$$

代入a得到

$$\begin{aligned}\int_{-\infty}^{\infty}y^{2n}\exp(-y^2/2\sigma^2)\mathrm{d}y&=(2\sigma^2)^{(2n+1)/2}(1/2)^n\sqrt{\pi}(2n)!/2^n n!\\&=\sigma^{2n}\sqrt{2\pi\sigma^2}(2n)!/2^n n!\end{aligned} \tag{5-299}$$

由于在整个表达式里，n总是伴随着2，可以定义偶数$m\triangleq 2n$。等式两边同除以$\sqrt{2\pi\sigma^2}$，定理得证。

因此，当n为偶数时，所有高斯分布的中心矩与方差成正比：

$$\mu_{c,n}=[n!/2^{n/2}(n/2)!]\mu_{c,2}^{n/2}\quad(\text{高斯}) \tag{5-300}$$

4阶中心矩有简单的表达式：$\mu_{c,4}=3\mu_{c,2}^2=3\sigma^4$。高斯随机变量的原点矩比较复杂，但当$\mu=0$时，它们可简化为式(5-291)。

定义(标准矩) 随机变量X的n阶**标准矩**等于n阶中心矩除以σ_X^n：

$$\tilde{\mu}_{c,n}\triangleq\mu_{c,n}/\sigma_X^n \tag{5-301}$$

分母中的指数确保标准矩无量纲。显然$\tilde{\mu}_{c,1}=0$、$\tilde{\mu}_{c,2}=1$。这里有两个重要概念，偏度$\tilde{\mu}_{c,3}$和峰度$\tilde{\mu}_{c,4}$，当定义完累积量后将讨论它们。

定义(累积量) 随机变量X的n阶累积量为：

$$\kappa_n\triangleq c_X^{(n)}(0) \tag{5-302}$$

其中$c_X^{(n)}(t)$是累积生成函数$c_X(t)$的n阶导。

本章稍后将定义累积生成函数。X的均值和方差分别是前两个累积量：$\mu_X\equiv\kappa_1=c_X^{(1)}(0)$、$\sigma_X^2\equiv\kappa_2=c_X^{(2)}(0)$。虽然这两个量分别是中心矩和原点矩，但高阶累积量一般与任何类型的矩都不完全一致，将更复杂。

偏度是分布不对称性的度量：累积分布函数之和或概率密度函数的面积集中在“中心”的左侧(偏度为正)或右侧(偏度为负)。对于单峰分布，下列关系通常适用于X的均值、

众数 m_o 和中位数 m_e，但也不总是这样。负偏度：$\mu_X < m_e < m_o$，正偏度：$m_o < m_e < \mu_X$。对于某些分布，上述表达式中的均值和中位数可能会互换位置。

定义(偏度)　随机变量 X 的**偏度**是：

$$\gamma_1 = \widetilde{\mu}_{c,3} \triangleq \mu_{c,3} / \sigma_X^3 \tag{5-303}$$

它也可以定义成与累积量有关：

$$\gamma_1 \triangleq \kappa_3 / \kappa_2^{3/2} \tag{5-304}$$

常用符号 γ_1 表示。

当概率密度函数对称时，均值、中位数和众数相等，如高斯分布，此时偏度为 0。从附录 A 分布的总结中发现，本书中讨论的大多数分布的偏度为正，只有少数分布的偏度为 0 或负值。

例 5-38　图 5-16 画出了三个概率密度函数以说明三种类型的偏度。对数正态分布的偏度为：

$$\gamma_1 = (\exp(\sigma^2) + 2)\sqrt{\exp(\sigma^2) - 1} \tag{5-305}$$

对于任意 $\sigma^2 > 0$，上式为正。二项式概率密度函数的偏度为：

$$\gamma_1 = (1 - 2p)/\sqrt{Npq} \tag{5-306}$$

当 $p \geq 1/2$ 时它为负值。对于任意 $\alpha > 0$，显然拉普拉斯分布的偏度为 0。

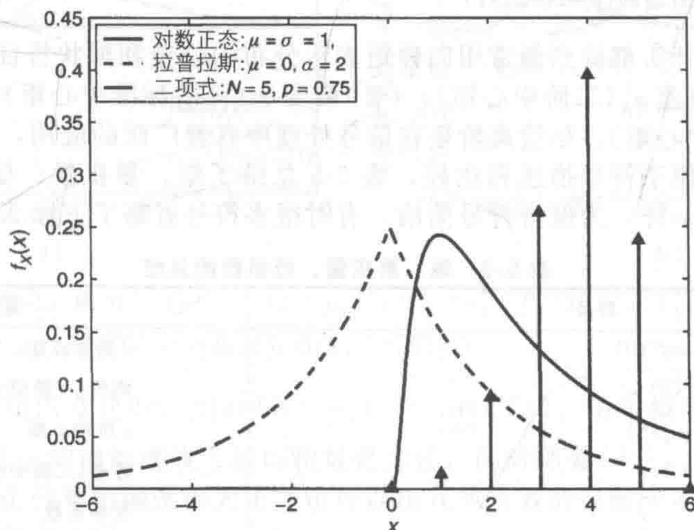


图 5-16　反映三种类型偏度的概率密度函数。(ⅰ)正 $\gamma_1 \approx 6.18$(对数正态)；(ⅱ)$\gamma_1 = 0$(拉普拉斯)；(ⅲ)负 $\gamma_1 \approx -1.19$(二项式)　◀

另一个用于描述分布特性的标准矩是峰度。

定义(峰度)　随机变量 X 的**峰度**为：

$$\widetilde{\mu}_{c,4} \triangleq \mu_{c,4} / \sigma_X^4 \tag{5-307}$$

其中 $\mu_{c,4}$ 是 4 阶中心矩。超额峰度是更为常见的形式，它由累积量定义：

$$\gamma_2 \triangleq \kappa_4 / \kappa_2^2 = \widetilde{\mu}_{c,4} - 3 \tag{5-308}$$

常用符号 γ_2 表示。

通常首选超额峰度，因为高斯随机变量的超额峰度为 0，因此“超额”是指相对于高斯分布而言。峰度是随机变量的概率密度函数相对于高斯概率密度函数尖锐或平坦的测度。但是从后面介绍的函数观点来看，峰度更多的作为分布重尾的测度(同样相对于高斯分布)。

例 5-39　图 5-17 比较了 3 个概率密度函数的峰度。瑞利概率密度函数的超额峰度为：

$$\gamma_2 = \frac{24\pi - 16 - 6\pi^2}{(4 - \pi)^2} \approx 0.2451 \tag{5-309}$$

这意味着它的重尾与高斯概率密度函数($\gamma_2 = 0$)的没有太大区别。另一方面，指数概率密度函数的 $\gamma_2 = 6$。第 3 章的表 3-4 总结了几种对称分布的尾部类型。

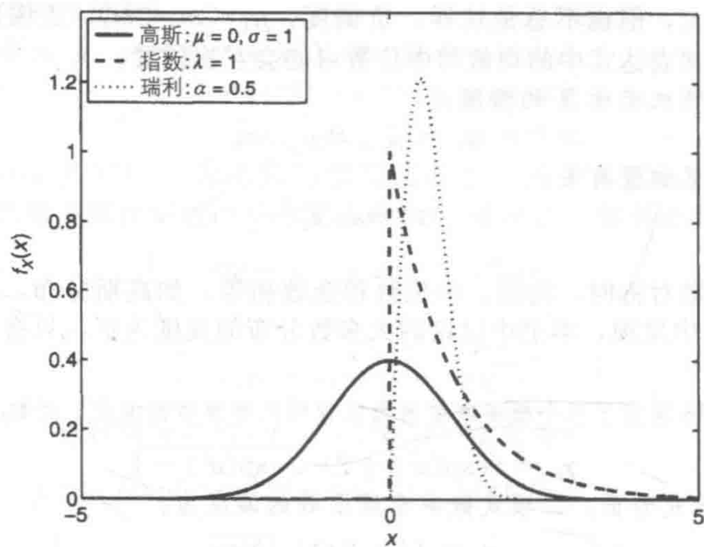


图 5-17　反映三种类型峰度的概率密度函数。(ⅰ)标准高斯($\gamma_2=0$);(ⅱ)$\lambda=1$ 的指数($\gamma_2=6$);(ⅲ)$\alpha=0.5$ 的瑞利($\gamma_2\approx 0.2451$)

第 3 章和本章开头都提到通常用四种矩表示分布的位置和形状特征:(ⅰ)均值 μ_X(一阶原点矩);(ⅱ)方差 σ_X^2(二阶中心矩);(ⅲ)偏度 γ_1(三阶标准中心矩);(ⅳ)超额峰度 γ_2(移位的四阶标准中心矩)。尽管高阶矩在信号处理中有着广泛的应用,但一般不使用高阶矩来描述分布。为便于符号描述和比较,表 5-5 总结了矩、累积量、母函数(稍后介绍)。与本书前面的做法一样,为保持符号简洁,有时很多符号省略了下标 X。

表 5-5　矩、累积量、母函数的总结

符号	定义
$\mu_n=\mathcal{E}[X^n]=m_X^{(n)}(0)$	n 阶原点矩
$\mu_X=\mu_1=M_X^{(n)}(0)$	均值(一阶原点矩)
$\mu_{c,n}=\mathcal{E}[(X-\mu_X)^n]$	n 阶中心矩
$\sigma_X^2=\mu_{c,2}$	方差(二阶中心矩)
$c_X=\sigma_X/\mu_X$	变异系数
$\widetilde{\mu}_{c,n}=\mu_{c,n}/\sigma_X^n$	n 阶标准中心矩
$\kappa_n=c_X^{(n)}(0)$	n 阶累积量
$\gamma_1=\widetilde{\mu}_{c,3}=\mu_{c,3}/\sigma_X^3=\kappa_3/\kappa_2^{3/2}$	偏度(三阶标准中心矩)
$\widetilde{\mu}_{c,4}=\mu_{c,4}/\sigma_X^4$	峰度(四阶标准中心矩)
$\gamma_2=\kappa_4/\kappa_2^2=\widetilde{\mu}_{c,4}-3$	超额峰度
生成函数	**类型**
$m_X(t)=\mathcal{E}[\exp(tX)]=\sum_{n=0}^{\infty}t^n\mu_n/n!$	矩母函数
$M_X(t)=\mathcal{E}[\exp(t(X-\mu_X))]=\sum_{n=0}^{\infty}t^n\mu_{c,n}/n!$	中心矩母函数
$c_X(t)=\ln(m_X(t))=\sum_{n=1}^{\infty}t^nK_n/n!$	累积量母函数
$\phi_X(\omega)=\mathcal{E}[\exp(\mathrm{j}\omega X)]=\sum_{n=1}^{\infty}(\mathrm{j}\omega)^n\mu_n/n!$	特征函数
$G_X(t)=\mathcal{E}[t^X]=\sum_{n=0}^{\infty}p_X[x]t^X$	概率母函数

5.22 偏度的函数观点

可以从函数的观点说明偏度是分布不对称性的测度。前面提过如果概率密度函数关于均值对称，则所有奇数阶中心距为0。这是因为当 n 为奇数时，$(X-\mu)^n$ 为奇函数，因此它的期望是零。奇数阶中心矩随着概率密度函数不对称性的增加而增大(始终是0的一阶中心矩 $\mathcal{E}[X-\mu]$ 除外)。三阶中心距为：

$$\mu_{c,3}=\mathcal{E}[(X-\mu_X)^3]=\int_{-\infty}^{\infty}(x-\mu_X)^3 f_X(x)\mathrm{d}x \tag{5-310}$$

图 5-18 画出了指数和瑞利概率密度函数 $(x-\mu_X)^3 f_X(x)$ 的例子，两个分布的方差均为1，因此 $\mu_{c,3}$ 和偏度相等。图中也画出了累积面积

$$C(x)=\int_{-\infty}^{x}(v-\mu_X)^3 f_X(v)\mathrm{d}v \tag{5-311}$$

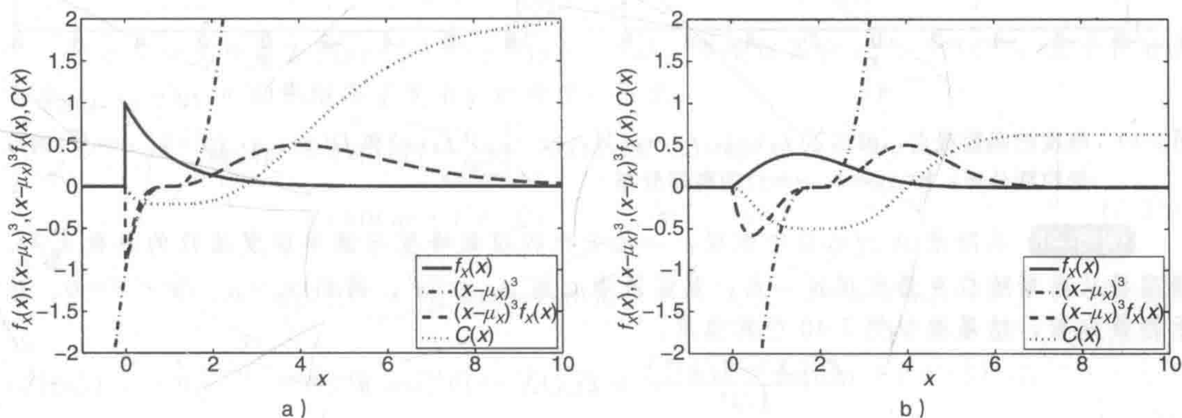


图 5-18 偏度的函数观点，画出了 $f_X(x)$，$(x-\mu_X)^3$，$(x-\mu_X)^3 f_X(x)$、$C(x)$。a) $\lambda=1$ 的指数分布($\mu_X=1$)；b) $\alpha=\sqrt{2/(4-\pi)}$ 的瑞利分布($\mu_X=1.9131$)

当 x 在 $f_X(x)$ 的区间内变化时，则 $\mathcal{E}[(X-\mu_X)^3]=\lim\limits_{x\to\infty}C(x)$。由于概率密度函数向左"倾斜"，观察发现随着 x 的增加虚线下的面积最终为正，同时随着 $(x-\mu_X)^3$ 的显著增加逐渐收敛到偏度。这两个分布的偏度都为正。也可以用其他奇数的高阶中心矩度量这种不对称性，但通常使用三阶中心矩。

例 5-40 这个例子给出了均值为 λ、方差为 $1/\lambda^2$ 的指数随机变量，对于任意 λ 值，其偏度为定值。根据定义：

$$\gamma_1=\frac{\mathcal{E}[(X-1/\lambda)^3]}{1/\lambda^3}=\mathcal{E}[(\lambda X-1)^3] \tag{5-312}$$

注意 $Y=\lambda X$ 为标准指数分布(均值为1)，因此上面的期望最终与 λ 无关。$\mathcal{E}[(Y-1)^3]=2$ 留作练习。从附录A的总结中可以看出，指数分布是本书少有的偏度与参数无关且是非零整数的概率密度函数之一。◀

5.23 峰度的函数观点

最后，从函数的观点说明峰度取决于概率密度函数的重尾。四阶中心矩为：

$$\mu_{c,4}=\mathcal{E}[(X-\mu_X)^4]=\int_{-\infty}^{\infty}(x-\mu_X)^4 f_X(x)\mathrm{d}x \tag{5-313}$$

图 5-19 画出了拉普拉斯和高斯概率密度函数的 $(x-\mu_X)^4 f_X(x)$ 和累积面积：

$$D(x)=\int_{-\infty}^{x}(v-\mu_X)^4 f_X(v)\mathrm{d}v \tag{5-314}$$

当 x 在 $f_X(x)$ 的区间内变化时，$\mathcal{E}[(X-\mu_X)^4]=\lim\limits_{x\to\infty}D(x)$。这两种情况下的方差均设置为

1，使得 $\mu_{c,4}$ 与峰度相等(减去 3 得到超额峰度)。观察发现拉普拉斯概率密度函数的虚线有更大的面积，从而其峰度比高斯概率密度函数的大。出现这种情况是因为exp($-x$)的尾部下降速度比 exp($-x^2$)慢，而$(x-\mu_X)^4$ 随着 x 大于或小于均值迅速增加。

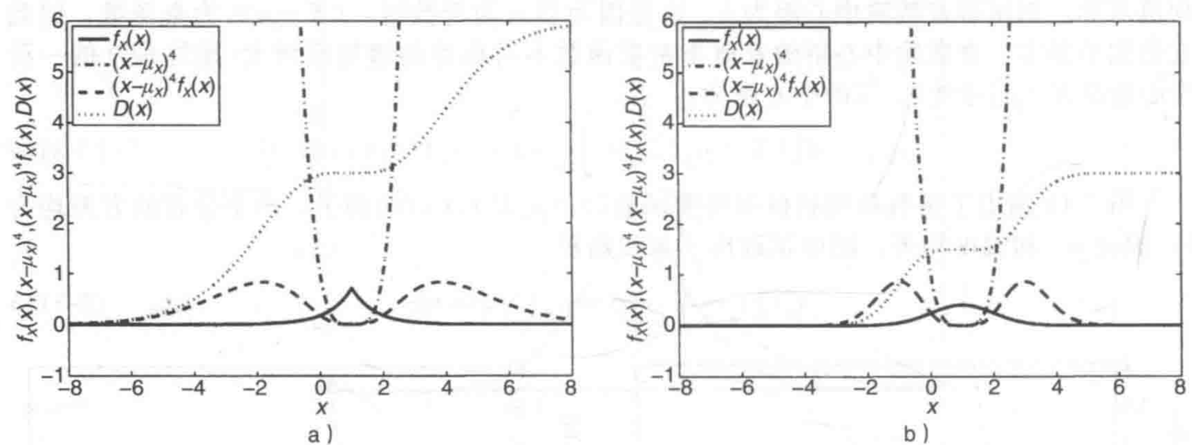


图 5-19 峰度的函数观点，画出了 $f_X(x)$，$(x-\mu_X)^4$，$(x-\mu_X)^4 f_X(x)$和 $D(x)$。a) $\{\mu=1, \alpha=\sqrt{2}\}$的拉普拉斯分布；b) $\{\mu=1, \sigma=1\}$的高斯分布

例 5-41 从附录 A 的总结中发现，一些分布的超额峰度与概率密度函数的参数无关。很容易从高斯随机变量发现这一点：其四阶中心矩 $\mu_{c,4}=3\sigma^2$，得到 $\gamma_2=\mu_{c,4}/\sigma^2-3=0$。对于指数分布，结果类似例 5-40 中的偏度：

$$\frac{\mathcal{E}[(X-1/\lambda)^4]}{1/\lambda^4}=\mathcal{E}[(\lambda X-1)^4]=9 \tag{5-315}$$

这是因为 λX 与 λ 无关。减去 3 得到 $\gamma_2=6$。最后，对于参数为 $\{\mu,\alpha\}$、方差为 $2\alpha^2$ 的拉普拉斯分布：

$$\frac{\mathcal{E}[(X-\mu)^4]}{(2\alpha^2)^2}=\frac{\mathcal{E}[Y^4]}{4\alpha^4}=\frac{1}{8\alpha^5}\int_{-\infty}^{\infty}y^4\exp(-|y|/2\alpha)\mathrm{d}y \tag{5-316}$$

其中 Y 是参数为$\{\mu=0, \alpha\}$的拉普拉斯分布。将变量改为 $z=y/\alpha$，得出：

$$\gamma_2=(1/8)\int_{-\infty}^{\infty}z^4\exp(-|z|/2)\mathrm{d}z-3=3 \tag{5-317}$$

对于那些峰度与比例参数无关的随机变量，显然四阶中心矩 $\mu_{c,4}$ 和分母的归一化系数 σ^4 之间有一定的关系。定理 5-11 明确给出了高斯随机变量两个参数间的关系。◀

5.24 母函数

本节将讨论像特征函数那样可以用于产生矩和累积量的期望。

定义(矩母函数) 随机变量 X 的**矩母函数**(moment generating function，MGF)为

$$m_X(t)\triangleq\mathcal{E}[\exp(tX)]=\sum_{n=0}^{\infty}\frac{t^n\,\mathcal{E}[x^n]}{n!}=\sum_{n=0}^{\infty}\frac{t^n\mu_n}{n!} \tag{5-318}$$

其中 $t\in\mathcal{R}$、$\mu_0=1$。n 阶矩生成如下：

$$\mu_n=m_X^{(n)}(0) \tag{5-319}$$

其中 $m_X^{(n)}(t)$是 $m_X(t)$关于 t 的 n 阶导。

矩母函数并不总是存在，但特征函数通常存在。回想一下，特征函数在指数$\Phi_X(\omega)=\mathcal{E}[\exp(\mathrm{j}\omega X)]$的指数项中包含 $\mathrm{j}=\sqrt{-1}$，且与傅里叶变换类似。很容易通过展开式(5-318)的最后一个表达式来验证 $m_X(t)$的矩母性质：

$$m_X(t)=1+t\mu_1+\frac{t^2\mu_2}{2!}+\frac{t^3\mu_3}{3!}+\cdots \tag{5-320}$$

例如，求导两次得到：

$$m_X^{(2)}(t) = 0 + 0 + \mu_2 + t\mu_3 + \cdots \tag{5-321}$$

显然，当 $t \neq 0$ 时，受 t 的幂次影响，$t\mu_3$ 及以后的各项均不为 0。令式(5-321)中的 t 等于 0，去掉这些项，仅保留了不是 t 的函数的二阶矩：

$$m_X^{(2)}(t)\big|_{t=0} = \mu_2 \tag{5-322}$$

例 5-42 设 X 是标准高斯随机变量。变换 $Y=\sigma X+\mu$ 产生了参数为 $\{\mu,\sigma\}$ 的另一个高斯随机变量。Y 的矩母函数是：

$$\begin{aligned} m_X(t) &= \mathcal{E}[\exp(t(\sigma X + \mu))] \\ &= \exp(t\mu)\int_{-\infty}^{\infty} \exp(t\sigma x)\,\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\mathrm{d}x \\ &= \exp(t\mu)\,\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp((2t\sigma x - x^2)/2)\mathrm{d}x \end{aligned} \tag{5-323}$$

写成平方的形式为 $2t\sigma x - x^2 = (x-t\sigma)^2 + t^2\sigma^2$。将该结果代入式(5-323)，由于积分是参数 $\{\mu=0,\sigma\}$ 的高斯概率密度函数的面积，因此：

$$\begin{aligned} m_X(t) &= \exp(t\mu)\,\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(t^2\sigma^2/2)\exp(-(x-t\sigma)^2/2)\mathrm{d}x \\ &= \exp(t\mu + t^2\sigma^2/2) \end{aligned} \tag{5-324}$$

从而：

$$m_X^{(1)}(0) = (\mu + t\sigma^2)\exp(t\mu + t^2\sigma^2/2)\big|_{t=0} = \mu = \mu_1 \tag{5-325}$$

且

$$\begin{aligned} m_X^{(2)}(0) &= \sigma^2\exp(t\mu + t^2\sigma^2/2)\big|_{t=0} + (\mu + t\sigma^2)^2\exp(t\mu + t^2\sigma^2/2)\big|_{t=0} \\ &= \sigma^2 + \mu^2 = \mu_2 \end{aligned} \tag{5-326}$$ ◀

矩母函数及其导数的例子如图 5-20 所示。根据图中导数在 $t=0$ 时刻的值得出矩。从标准高斯随机变量(均值为 0)对应的图 5-42a 可以得出结论，由于奇数阶导数必然是奇函数，因此奇数阶矩为 0。由于 $\mu=0$，图 5-42a 中的原点矩满足高斯中心矩的计算式(5-291)。

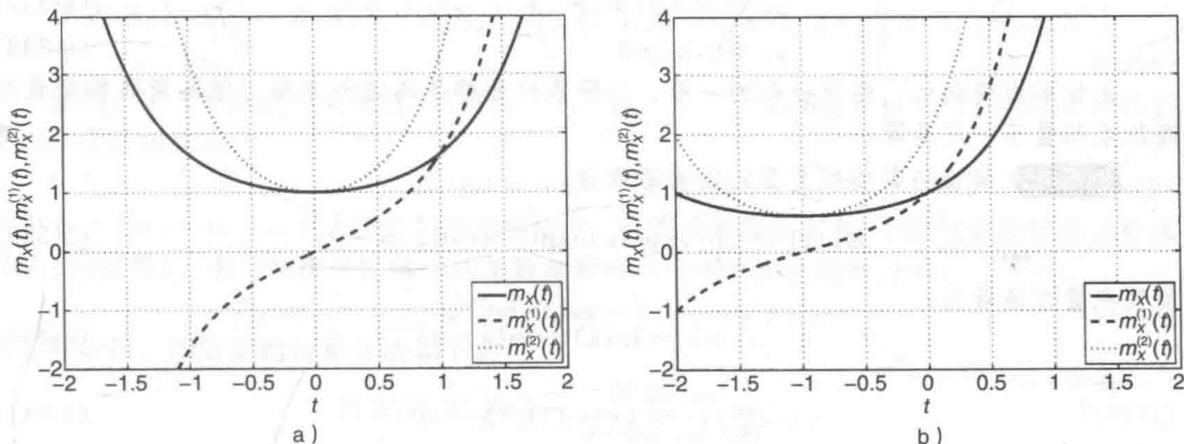


图 5-20 例 5-42 的高斯矩母函数及其导数。a) $\mu=0$，$\sigma=1$；b) $\mu=1$，$\sigma=1$

例 5-43 设 X 是参数为 $\{N,p\}$ $(q \triangleq 1-p)$ 的二项式随机变量。它的矩母函数是：

$$\begin{aligned} m_X(t) &= \sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}\exp(tn) \\ &= \sum_{n=0}^{N}\binom{N}{n}[p\exp(t)]^n q^{N-n} = [p\exp(t) + q]^N \end{aligned} \tag{5-327}$$

其中使用了附录 E 中的二项式公式。观察发现

$$m_X^{(1)}(0) = N[p\exp(t)+q]^{N-1}p\exp(t)\big|_{t=0} = Np = \mu_1 \tag{5-328}$$

且

$$\begin{aligned} m_X^{(2)}(0) &= N(N-1)[p\exp(t)+q]^{N-2}p^2\exp(t)\big|_{t=0} + N[p\exp(t)+q]^{N-1}p\exp(t)\big|_{t=0} \\ &= p^2N(N-1)+Np = N^2p^2+Npq = \mu_2 \end{aligned} \tag{5-329}$$

下面的矩母函数的扩展也许不是众所周知的。 ◀

定义（中心矩母函数） 随机变量 X 的**中心矩母函数**(central moment generating function, CMGF)是

$$M_X(t) \triangleq \mathcal{E}[\exp(t(X-\mu_X))] = \sum_{n=0}^{\infty}\frac{t^n\,\mathcal{E}[(X-\mu_X)^n]}{n!} = \sum_{n=0}^{\infty}\frac{t^n\mu_{c,n}}{n!} \tag{5-330}$$

其中 $t\in\mathcal{R}$、$\mu_{c,0}=1$。N 阶中心矩产生如下：

$$\mu_{c,n} = M_X^{(n)}(0) \tag{5-331}$$

中心矩母函数可以写成矩母函数的形式：

$$M_X(t) = \exp(-\mu_X t)\,\mathcal{E}[\exp(tX)] = \exp(-\mu_X t)m_X(t) \tag{5-332}$$

下列母函数可用于产生累积量。

定义（累积量母函数） 随机变量 X 的**累积量母函数**(cumulant generating function, CGF)是矩母函数的对数：

$$c_X(t) \triangleq \ln(\mathcal{E}[\exp(tX)]) = \ln(m_X(t)) = \sum_{n=1}^{\infty}\frac{t^n\kappa_n}{n!} \tag{5-333}$$

对 $c_X(t)$ 关于 t 求 m 阶导并令 $t=0$，得到 m 阶累积量：

$$\kappa_n = c_X^{(m)}(0) \tag{5-334}$$

需要注意求和的下限是 1(受对数影响)，可认为 $n=0$ 时该项为 0。

例 5-44 例 5-42 中高斯随机变量矩母函数的对数是：

$$c_X(t) = t\mu + t^2\sigma^2/2 \tag{5-335}$$

很容易得到累积量如下：

$$\kappa_1 = c_X^{(1)}(0) = \mu + t\sigma^2\big|_{t=0} = \mu \tag{5-336}$$

$$\kappa_2 = c_X^{(2)}(0) = \sigma^2 \tag{5-337}$$

$$\kappa_n = 0, n\geqslant 3 \tag{5-338}$$

正如前面提到的，任何分布的一阶、二阶累积量都是均值和方差，但高斯随机变量的高阶累积量不一定是零。 ◀

例 5-45 假设指数随机变量的矩母函数为：

$$m_X(t) = \lambda\int_0^{\infty}\exp(tx)\exp(-\lambda x)\,\mathrm{d}x = \frac{\lambda}{\lambda-t} \tag{5-339}$$

其累积量母函数为：

$$c_X(t) = \ln(\lambda) - \ln(\lambda - t) \tag{5-340}$$

由于：

$$c_X^{(n)}(t) = (n-1)!/(\lambda-t)^n \tag{5-341}$$

累积量有以下简单的形式：

$$\kappa_n = (n-1)!/\lambda^n \tag{5-342}$$

最后，提出一个只适用于离散随机变量的母函数。 ◀

定义（概率母函数） 离散随机变量 X 的**概率母函数**(probability generating function, PGF)为：

$$G_X(t) \triangleq \mathcal{E}[t^X] = \sum_{x=0}^{\infty}p_X[x]t^x \tag{5-343}$$

$t\in\mathcal{R}$。对 $G_X(t)$ 关于 t 求 n 阶导并令 $t=0$，得到概率为：

$$P(X=m)=G_X^{(m)}(0)/m! \tag{5-344}$$

由于求 m 阶导后的第一个非0项为 $m!\,p_X[m]$，所以最终的结果在分母里有 $m!$。很容易验证下列性质：

- 概率和　$G_X(1)=1$（即 $p_X(x)$ 是有效的概率质量函数）。
- 特征函数(CF)　$\Phi_X(\mathrm{j}\omega)=G_X(\exp(\mathrm{j}\omega))$。
- 期望　$\mathcal{E}[X]=G_X^{(1)}(1)$。
- 矩母函数(MGF)　$m_X(t)=G_X(\exp(t))$。
- 阶乘矩母函数(FMGF)　$\mathcal{E}[(X)_n]=G_X^{(n)}(1)$，其中 $(X)_n$ 是递减的阶乘（见附录B）。
- 方差　$\sigma_X^2=G_X^{(2)}(1)+G_X^{(1)}(1)[1-G_X^{(1)}(1)]$。
- X 和 Y 独立　$G_{X+Y}(t)=G_X(t)G_Y(t)$。

需要注意的是，如果 $t=1$ 时期望存在，用于阶乘矩母函数的 $G_X(t)$ 也适用于连续随机变量。

例 5-46 对于几何随机变量：

$$G_X(t)=\sum_{x=0}^{\infty}pq^xt^x=\frac{p}{1-qt} \tag{5-345}$$

这里 $G(1)=1$，则

$$\mathcal{E}[X]=\frac{pq}{(1-qt)^2}\Big|_{t=1}=q/p \tag{5-346}$$

$$P(X=1)=\frac{pq}{(1-qt)^2}\Big|_{t=0}=pq \tag{5-347}$$ ◀

5.25 高斯四阶矩

下面的定理给出了高斯随机变量的四阶矩，在后面第12章自适应滤波器中将用到它。

定理 5-12 设 $\{X_1,X_2,X_3,X_4\}$ 是零均值联合高斯随机变量。它们的联合矩为：

$$\mathcal{E}[X_1X_2X_3X_4]=\mathcal{E}[X_1X_2]\mathcal{E}[X_3X_4]+\mathcal{E}[X_1X_3]\mathcal{E}[X_2X_4]+\mathcal{E}[X_1X_4]\mathcal{E}[X_2X_3] \tag{5-348}$$

证明： 定义高斯随机向量 $\boldsymbol{X}\triangleq[X_1,\ X_2,\ X_3,\ X_4]^{\mathrm{T}}$，将 $\boldsymbol{R_{XX}}$ 和 $\boldsymbol{\mu}=0$ 代入式(5-109)后得出其特征函数为：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=\exp(-\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{XX}}^{-1}\boldsymbol{\omega}) \tag{5-349}$$

其中 $\boldsymbol{\omega}=[\omega_1,\omega_2,\omega_3,\omega_4]^{\mathrm{T}}$。由于 $\boldsymbol{X}$ 的均值为0，则该函数值是实数（不像大多数特征函数那样是 j 的函数）。通常情况下，这样的指数函数可以用麦克劳林级数分析：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=1+(1/2)\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{R_{XX}}\boldsymbol{\omega}+(1/8)(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{R_{XX}}\boldsymbol{\omega})^2+\cdots \tag{5-350}$$

表示高阶项。四阶矩的计算方法如下：

$$\mathcal{E}[X_1X_2X_3X_4]=\frac{\partial^4\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})}{\partial\omega_1\partial\omega_2\partial\omega_3\partial\omega_4}\Big|_{\boldsymbol{\omega}=0} \tag{5-351}$$

求导并将 $\boldsymbol{\omega}=0$ 代入后，式(5-350)中只有第三项非0：

$$\mathcal{E}[X_1X_2X_3X_4]=(1/8)\frac{\partial^4(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{R_{XX}}\boldsymbol{\omega})^2}{\partial\omega_1\partial\omega_2\partial\omega_3\partial\omega_4}\Big|_{\boldsymbol{\omega}=0} \tag{5-352}$$

自相关矩阵 $\boldsymbol{R_{XX}}$ 的16个元素均为 $\mathcal{E}[X_mX_n]$ 的形式，其中 $m,\ n=1,\ 2,\ 3,\ 4$，由下式得出 $(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{R_{XX}}\boldsymbol{\omega})^2$ 有 $16^2=256$ 项，

$$(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{R_{XX}}\boldsymbol{\omega})^2=\sum_{k=1}^{4}\sum_{l=1}^{4}\sum_{m=1}^{4}\sum_{n=1}^{4}\omega_k\omega_l\omega_m\omega_n\,\mathcal{E}[X_kX_l]\mathcal{E}[X_mX_n] \tag{5-353}$$

然而，由于微分后代入了$\boldsymbol{\omega}=0$，因此只保留了下标$\{\omega_k,\omega_l,\omega_m,\omega_n\}$互不相同的项，因为其他项有一个或多个$\omega$的次数超过1。虽然有4！=24个这样的非零项，但互不相同的只有3项，每项出现了8次。因此，消去式(5-352)的因子1/8，最终：

$$\mathcal{E}[X_1X_2X_3X_4]=\mathcal{E}[X_1X_2]\mathcal{E}[X_3X_4]+\mathcal{E}[X_1X_3]\mathcal{E}[X_2X_4]+\mathcal{E}[X_1X_4]\mathcal{E}[X_2X_3] \tag{5-354}$$

可以看出，零均值高斯随机变量的四阶矩可简化为二阶矩之和。事实上，所有零均值联合高斯随机变量的偶数阶矩都可以简化为二阶原点矩的函数，而其奇数阶矩都是零(类似前面的一维高斯随机变量)。当$\{X_n\}$的方差均为σ^2时(或当它们是同一随机变量X的样本时)，式(5-348)简化为：

$$\mathcal{E}[X_1X_2X_3X_4]=3\sigma^2 \tag{5-355}$$

这是式(5-291)中的$n=2$，均值为0的结果。

5.26 非线性变换的期望

一般来说，很难得到几个随机变量非线性函数期望的闭合表达式。但是，如果随机变量服从互相关非零的联合高斯分布，当非线性函数无记忆时，有可能得到有用的期望。由于无记忆非线性经常出现在工程问题中，而高斯噪声总是不可避免地存在，因此分析该期望有着重要的现实意义。

下面的定理将互相关函数的导数和非线性函数的导数联系在一起。因此，如果非线性函数的导数是一个简单的表达式，可以很容易计算出期望。尽管该定理适用于N维联合高斯随机变量，但这里仅给出$N=2$的结果，并且没有给出证明过程(Price，1958；McMahon，1964；Pawula，1967)。

定理5-13(Price，et al) 设X_1和X_2是相关系数为ρ的联合高斯随机变量，并假设它们通过了无记忆非线性函数：$g_1(X_1)$和$g_2(X_2)$。定义变换后的随机变量的互相关函数为：

$$R_g\triangleq\mathcal{E}[g_1(X_1)g_2(X_2)] \tag{5-356}$$

那么

$$\frac{\partial^n R_g}{\partial\rho^n}=\mathcal{E}\left[\frac{\partial^n g_1(X_1)}{\partial X_1^n}\frac{\partial^n g_2(X_2)}{\partial X_2^n}\right] \tag{5-357}$$

对于更一般的函数$h(X_1,X_2)$，互相关为：

$$R_h\triangleq\mathcal{E}[h(X_1,X_2)] \tag{5-358}$$

微分为：

$$\frac{\partial^n R_h}{\partial\rho^n}=\mathcal{E}\left[\frac{\partial^{2n}h(X_1,X_2)}{\partial X_1^n\partial X_2^n}\right] \tag{5-359}$$

例5-47 考虑“硬限制”非线性符号函数：

$$g_1(x)=g_2(x)=\operatorname{sgn}(x) \tag{5-360}$$

一阶导是狄拉克δ函数：

$$\frac{\partial}{\partial x}g_m(x)=2\delta(x) \tag{5-361}$$

假设零均值随机变量$\{X_1,X_2\}$具有单位方差。将式(5-361)代入式(5-357)得到

$$\begin{aligned}\frac{\partial R_g}{\partial\rho}&=4\,\mathcal{E}[\delta(X_1)\delta(X_2)]\\&=\frac{4}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x_1)\delta(x_2)\exp(-[x_1^2-2\rho x_1x_2+x_2^2]/2(1-\rho^2))\mathrm{d}x_1\mathrm{d}x_2\\&=\frac{2}{\pi\sqrt{1-\rho^2}}\end{aligned} \tag{5-362}$$

这里利用了二元高斯概率密度函数。观察发现δ函数大大简化了二重积分，它的筛选特性导致指数函数只在$x_1=x_2=0$有非零值。这个例子说明了该定理的用处：对某些非线性函数进行微分(特别是包含阶跃或符号函数时)后，可能产生期望的简化形式。最后，对式(5-362)关于ρ进行积分可以得到R_g，但需要确定积分下限。当$\rho=0$时，$\{X_1, X_2\}$不相关，式(5-356)的期望分成两部分，由于非线性函数为奇函数，因此各部分均为0，从而$R_g=0$。所以

$$\int_0^{R_g} \mathrm{d}R_g = \int_0^{\rho} \frac{2}{\pi\sqrt{1-\rho^2}}\mathrm{d}p \tag{5-363}$$

$$R_g = \mathcal{E}[\operatorname{sgn}(X_1)\operatorname{sgn}(X_2)] = (2/\pi)\arcsin(\rho) \tag{5-364}$$

即为要求的期望。图5-21中画出了该结果，横坐标为相关系数ρ。受符号函数的影响，$|R_g|\leqslant 1$。◀

例5-48 保持例5-47随机变量的条件不变，考虑更一般的非线性函数：

$$h(x_1,x_2) = |x_1+x_2| - |x_1-x_2| \tag{5-365}$$

其对应于一个"整流器"。由于该函数不能分成两个函数的乘积，所以使用式(5-359)对应的定理的扩展。微分结果为：

$$\frac{\partial h(x_1,x_2)}{\partial x_1} = \operatorname{sgn}(x_1+x_2) - \operatorname{sgn}(x_1-x_2) \tag{5-366}$$

$$\frac{\partial^2 h(x_1,x_2)}{\partial x_1 \partial x_2} = 2[\delta(x_1+x_2) - \delta(x_1-x_2)] \tag{5-367}$$

δ函数再次大大简化了期望：

$$\begin{aligned}\mathcal{E}\left[\frac{\partial^2 h(X_1,X_2)}{\partial X_1 \partial X_2}\right] &= \frac{1}{\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\delta(x_1+x_2)-\delta(x_1-x_2)] \\ &\quad \exp(-(x_1^2-2\rho x_1x_2+x_2^2)/2(1-p^2))\mathrm{d}x_1\mathrm{d}x_2 \\ &= \frac{1}{\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}[\exp(-x_1^2/(1+\rho))+\exp(-x_1^2/(1-\rho))]\mathrm{d}x_1\end{aligned} \tag{5-368}$$

被积函数包括两个零均值高斯概率密度函数，其方差分别为：$(1-\rho)/2$和$(1+\rho)/2$。利用方差为σ^2的高斯概率密度函数的积分等于：

$$\int_{-\infty}^{\infty} \exp(-x^2/2\sigma^2)\mathrm{d}x = \sqrt{2\pi\sigma^2} \tag{5-369}$$

得到：

$$\mathcal{E}\left[\frac{\partial^2 h(X_1,X_2)}{\partial X_1 \partial X_2}\right] = \frac{\sqrt{\pi(1+\rho)}+\sqrt{\pi(1-\rho)}}{\pi\sqrt{1-\rho^2}} = \frac{1}{\sqrt{\pi}}\left[\frac{1}{\sqrt{1-\rho}}+\frac{1}{\sqrt{1+\rho}}\right] \tag{5-370}$$

最后，利用边界条件$R_h=0$时$\rho=0$，式(5-370)关于ρ进行积分如下：

$$\int_0^{R_h}\mathrm{d}R_h = \frac{1}{\sqrt{\pi}}\int_0^{\rho}\left[\frac{1}{\sqrt{1-\rho}}+\frac{1}{\sqrt{1+\rho}}\right]\mathrm{d}\rho \tag{5-371}$$

得到期望：

$$R_h = \mathcal{E}[|X_1+X_2| - |X_1-X_2|] = \frac{2}{\sqrt{\pi}}[\sqrt{1+p}-\sqrt{1-p}] \tag{5-372}$$

此结果也绘制在图5-21中。有趣的是，当随机变量的$|\rho|=1$也就是完全相关时，$|R_h|\approx 1.5958$，该值小于$X_1=\pm X_2$时对应的$|R_h|=2$。这个例子表明，随机变量完全相关和相等(概率为1)不等价。

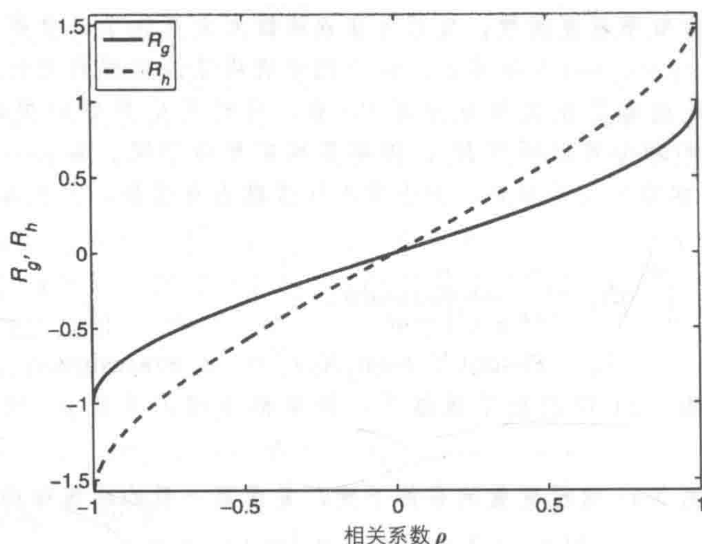


图 5-21　分别利用式(5-364)和式(5-372)得到的例 5-47 和例 5-48 相关系数 ρ 对应的非线性期望 ◀

习题

简单随机变量

5-1　令 X 是参数 $\{\mu=0,\alpha\}$ 的拉普拉斯随机变量。定义映射：

$$Y \triangleq \begin{cases} 1, & |X| \leqslant 1 \\ 1/2, & |X| > 1 \end{cases} \tag{5-373}$$

写出简单随机变量 Y 的表达式并给出它的概率质量函数。

5-2　证明二项随机变量是一个简单随机变量。

5-3　将实线 $\Omega=\mathcal{R}$ 分成以下三个区间：$(-\infty,0]$，$(0,2]$ 和 $(2,\infty)$ 定义

$$Y \triangleq \begin{cases} -1, & X \leqslant 0 \\ 0, & 0 < X \leqslant 2 \\ 1, & X > 2 \end{cases} \tag{5-374}$$

其中 X 是标准高斯随机变量。写出简单随机变量 Y 的表达式并给出它的概率质量函数。

期望

5-4　(a)证明如果 X 是 $\mathcal{E}[X]=0$ 的非负离散随机变量，那么 $P(X=0)=1$。
(b)重复(a)，如果 $\mathcal{E}[X]=1$，那么 $P(X=1)=1$。

5-5　计算参数 $\{N,p\}$ 的二项随机变量的 $\mathcal{E}[X]$（不能利用特征函数和矩母函数）。

5-6　对于参数 $\{N,M,n\}$ 的超几何随机变量，重复习题 5-5。

5-7　证明非负离散随机变量 X 的 $\mathcal{E}[X]=\sum_{n=1}^{\infty}P(X\geqslant n)$。

5-8　如果连续随机变量 X 的期望有限，证明 $\lim_{x\to\infty}|x|f_X(x)=0$。

5-9　计算参数 $\{\mu,\sigma\}$ 的对数正态分布随机变量的 $\mathcal{E}[X]$（不使用特征函数和矩母函数）。

5-10　对于参数 $\{\alpha,\beta\}$ 的贝塔随机变量，重复习题 5-9。

函数的期望

5-11　设 X 是参数 $\{\mu=0,\ \sigma\}$ 的高斯随机变量。(a)求 $Y=u(X)$ 的 $\mathcal{E}[Y]$ 和方差 σ_Y^2，其中 $u(\cdot)$ 是单位阶跃函数；(b)对于 $Z=u(X-1)$，重复(a)。

5-12　利用习题 5-11 中的 X，求 $Y=|X|$ 的 $\mathcal{E}[Y]$ 和方差 σ_Y^2（Y 为半正态分布）。

5-13　求 $Y=\text{sgn}(X-5)$ 的期望，其中 X 是参数为 p 的几何随机变量，取值范围为 $\mathcal{Z}^+$。

5-14　求 $Y=\exp(|X-1/2|)$ 的期望，其中 X 是 $[-1,\ 1]$ 的均匀分布。

5-15　将 $\mathcal{E}[u(X-c)]$ 写成累积分布函数 $F_X(x)$ 的形式，其中 $u(\cdot)$ 是单位阶跃函数。利用(a)期望的定义；(b)条件期望；(c)对于 $\mathcal{E}[\text{sgn}(X-c)]$ 重复上述问题。

5-16　设 X 是参数 $\lambda>1$ 的指数随机变量。求(a) $\mathcal{E}[\exp(X)]$；(b) $\mathcal{E}[(1-X)I_{[0,1]}(X)]$，其中 $I(\cdot)$ 是指示函数。

5-17　设 $\{X_1,\cdots,X_N\}$ 是参数为 λ 的独立同分布的指数随机变量。求 $\mathcal{E}[X_{(1)}]$ 和 $\mathcal{E}[X_{(N)}]$，其中 $X_{(1)}$ 和 $X_{(N)}$ 是最小值和最大值（顺序统计量）。

5-18　利用特征函数证明，当 X 和 Y 相互独立时，$Z=X-Y$ 的概率密度函数是确定性互相关 $f_Z(z)=f_Y(y)*f_X(x)$（见式(4-160)）。

条件期望

5-19　设 X 和 Y 有三角二元概率密度函数 $f_{X,Y}(x,\ y)=[u(y)-u(y-1)][u(x+y-1)-u(x-y+1)]$。求 $\mathcal{E}[XY]$ 和 $\mathcal{E}[Y|X]$。

5-20　当 $\{X_n\}$ 独立同分布并且与 N 相互独立时，随机和 $Y=\sum_{n=1}^{N}X_n$ 的均值是 $\mathcal{E}[Y]=\mathcal{E}[X]\mathcal{E}[N]$。求 Y 在条件 N 下的方差。

5-21 求参数 $\{N,p\}$ 的二项随机变量 X 的$\mathcal{E}[X|X>2]$。

5-22 设 $Y=\prod_{n=1}^{N}X_n$，其中$\{X_n\}$是独立同分布且在[0，1]均匀分布的随机变量，N 是参数为 α 的泊松分布。假定$\{X_n\}$和 N 独立，求条件 N 下的$\mathcal{E}[Y]$。

5-23 拉普拉斯随机变量的参数为 $\{\mu=0,\alpha\}$，求 (a)$\mathcal{E}[X||X|<1]$；(b)$\mathcal{E}[X||X<1]$。

5-24 设 X 和 Y 是独立同分布的在[−1，1]上均匀分布的随机变量。求(a)$\mathcal{E}[XY|X>Y]$；(b)$\mathcal{E}[X-Y|X>Y]$

5-25 对在 $\{1,\cdots,N\}$ 上均匀分布的离散随机变量，重复习题 5-24。

位置参数

5-26 计算参数为 α 的瑞利分布的均值、中位数和众数。

5-27 计算参数为 $\{c,\alpha\}$ 的极值(康拜尔)分布的中位数和众数。

5-28 对于参数为 $\{\mu,\sigma\}$ 的对数正态分布，重做习题 5-27。(均值可直接利用习题 5-9 的结论)。

5-29 证明非对称拉普拉斯分布的均值为式(3-106)，并写出中位数 m_e 的表达式。

方差、协方差和相关

5-30 (a)利用期望的定义，计算参数为$\{\alpha,r\}$的埃尔朗随机变量的均值和方差；(b)利用指数随机变量的结果，重复(a)。

5-31 利用协方差定义，计算下列联合概率密度函数对应的 C_{XY}，其中 $\{a,b\}$ 是常数。

(a) $f_{X,Y}(x,y)=axy[u(x)-u(x-1)][u(y)-u(y-1)]$

(b) $f_{X,Y}(x,y)=ax\exp(-b(x+y))u(x)u(y)$

5-32 计算几何随机变量的$\mathcal{E}[X^2]$和 σ_X^2。(a)区间为$\mathcal{Z}^+$(b)区间为$\mathcal{N}$。

5-33 导出式(5-242)中的皮尔逊样本相关系数的表达式。

5-34 根据随机变量 X 和$\hat{Y}=a_1X+a_2X^2$ 计算 Y 的下列估计：(a)使$\mathcal{E}[(Y-\hat{Y})^2]$最小的系数$\{a_1,a_2\}$的表达式；(b)当 X 和 Y 服从均值为 0，方差为 1，相关系数为 ρ_{XY} 的联合高斯分布时，重新计算满足条件的 $\{a_1,a_2\}$。

5-35 导出二元指数分布的相关系数为式(4-86)。

5-36 计算下列随机变量的方差：(a)在$\{1,\cdots,N\}$均匀分布的离散随机变量；(b)参数为 p 的对数分布随机变量。

5-37 推导式(5-261)中累积分布函数对应的联合概率密度函数 $f_{X,Y}(x,y)$。

母函数

5-38 (a)求参数为 α 的泊松随机变量的矩母函数 $m_X(t)$；(b)求参数为$\{N,p\}$的二项式随机变量的特征函数 $\Phi_X(j\omega)$。

5-39 求下列随机变量的特征函数：(a)参数为$\{\mu,\alpha\}$ 的拉普拉斯随机变量；(b)参数为$\{\lambda,r\}$的埃尔朗随机变量。

5-40 证明特征函数的上限为下式：

$$|\Phi_X(\omega)|\leqslant\Phi_X(0)=1 \tag{5-375}$$

5-41 利用特征函数计算$\mathcal{E}[X^2]$：(a)X 是参数为 α 的泊松随机变量；(b)Y 是参数为$\{\mu=0,\alpha\}$的拉普拉斯随机变量。

5-42 证明多元高斯随机变量的特征函数为式(5-109)。

矩和累积量

5-43 对于参数为 p 的几何随机变量，计算$\mathcal{E}[X(X-1)\cdots(X-n)]$的表达式。

5-44 导出在$[a,b]$上均匀分布的连续随机变量的 n 阶中心距 $\mu_{c,n}$。

5-45 计算参数为 α 的泊松随机变量的所有累积量。

5-46 将三阶累积量 κ_3 的表达式写成原点矩 $\mu_n=\mathcal{E}[X^n]$的形式。其中 $n=1,2,3$。

非线性函数的期望

5-47 假设$\{X_1,X_2\}$是均值为 0、相关系数为 ρ 的联合高斯随机变量，求$\mathcal{E}[u(X_1)u(X_2-1)]$的表达式。

5-48 保持习题 5-47 的条件不变，求$\mathcal{E}[X_1u(X_1)X_2u(X_2-1)]$的表达式。

仿真作业

5-49 设 $\hat{Y}=a_1+a_2X$ 是 Y 的估计值，求使$\mathcal{E}[(Y-\hat{Y})^2]$最小的最佳系数$\{a_1,a_2\}$。利用 mvnrnd 产生 1000 个样本$\{X,Y\}$，其中 $\mu_X=1$，$\mu_Y=2$，$\sigma_X^2=\sigma_Y^2=1$，$\rho_{XY}=0.8$。画出样本的散点图，并在上述参数的最佳系数的基础上绘制直线。保持其他条件不变，令 $\rho_{XY}=-0.8$，重新画图。

5-50 使用 mvnrnd 产生 1000 个样本$\{X,Y\}$，其中 $\mu_X=0$、$\mu_Y=1$、$\sigma_X^2=1$，$\sigma_Y^2=2$，$\rho_{XY}=0.9$。画出样本的散点图，并使用 MATLAB 计算：(a)皮尔逊(Pearson)样本相关系数 $\bar{\rho}_{XY}$；(b)斯皮尔曼(Spearman)等级相关系数 τ_S；(c)肯德尔(Kendall)等级相关系数 τ_K。

5-51 对于式(5-118)的随机和，假设 N 服从参数$p=0.1$ 的几何分布，$\{X_n\}$是参数 $\lambda=1$ 的独立同分布指数随机变量。从条件期望知道，当 N 和$\{X_n\}$独立时，$\mathcal{E}[Y]=\mathcal{E}[\mathcal{E}[Y|N]]=\mathcal{E}[X_n]\mathcal{E}[N]=(1/\lambda)q/p=9$。利用 MATLAB 产生随机变量 N 的结果 n，然后求 n 个指数样本的平均来估计$\mathcal{E}[Y|N]$。对 1000 个 N 的样本重复上述步骤，计算样本均值 $\bar{Y}$ 用于估计$\mathcal{E}[Y]$。

进一步阅读

第 2～4 章的参考文献里很多包含期望的其他资料。附录 D 的结尾给出了积分的部分参考文献。

第二部分

随机过程、系统与参数估计



第6章 随机过程

6.1 引言

随机过程是随着时间而改变的随机变量的集合。设随机过程定义在时间集合$\mathcal{T}$上。这个集合可以是连续时间，比如$\mathcal{T}=\mathcal{R}$或$\mathcal{R}^+$，也可以是离散时间$\mathcal{T}=\mathcal{Z}$或$\mathcal{Z}^+$。离散时间随机过程也被称为随机序列或时间序列，这些离散时刻通常都是等分的，尽管不一定是必需的。在本章中，我们经常用一般意义下随机过程这一术语，包括了连续时间随机过程和离散时间序列。当必须把讨论限制在离散时间时，我们使用随机序列这一术语。随机过程通常也称为"stochastic process"(stochastic 来源于希腊语 stokhastikos，意思是"猜的能力")。

图 6-1 显示的是一个高斯随机过程的现实。我们感兴趣的是探索随机过程的各种特征，这有助于我们理解随机过程的性质。这个时间函数是通过对零均值高斯随机变量组成的独立序列滤波得到的。因此与几个时间点对应的随机变量是联合高斯的。我们也知道这个滤波器使得输出为一个相关序列，因此输出在某种程度上是可预测的。本章中，我们将描述随机过程的几个其他特征，包括：平稳性、独立增量、鞅和马尔可夫特性。此外还将讨论几个特殊的随机过程，重点讨论他们的相关特性。在第 7 章，我们还将学习随机过程的其他性质，如随机收敛、连续性、导数和积分。随机微分方程的学习是为第 8 章中描述的系统做准备，系统通过对随机过程滤波改变其特性，就如上文中提到的高斯随机过程的例子。

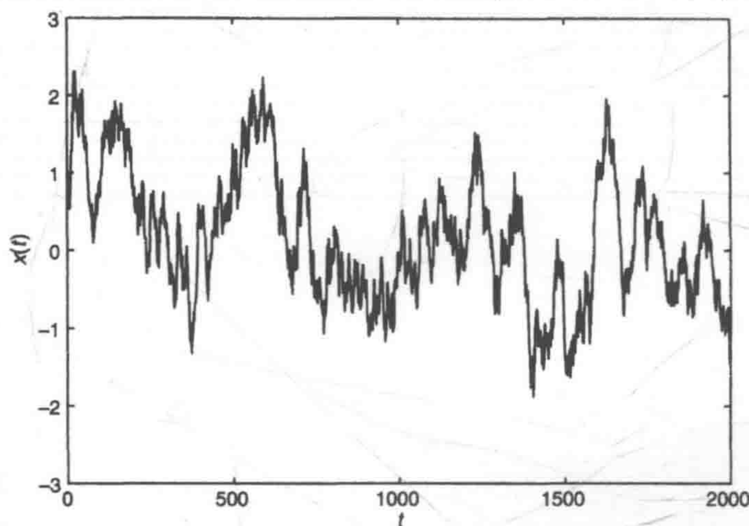


图 6-1 相关高斯随机过程的现实

6.2 随机过程的特征

随机过程可按与随机变量类似的方法定义。假设存在一个样本空间$\mathcal{S}$，包括随机过程在$\mathcal{T}$内所有可能的试验结果。我们用 $\zeta\in\mathcal{S}$表示一个试验结果，用这个符号来表示随机过程的一个试验结果是基于以下两个原因：(ⅰ)为了与之前用来表示一个抽象样本空间的试验结果 $\omega\in\Omega$ 区分开来；(ⅱ)避免与正弦信号和傅里叶变换的角频率 ω 发生混淆。$\mathcal{S}$的每一个试验结果都是一个关于时间的确定函数，称为随机过程的一个现实，我们假设用一个基本概率空间 $\{S,\mathcal{F},P\}$ 来定义这个随机过程。对应于试验结果 $\zeta\in\mathcal{S}, x(t,\zeta)$ 表示这个现实。

由所有可能的试验结果组成的现实的总体，用大写的函数 $X(t,\zeta)$ 来表示。图 6-2 就是两个连续时间的现实的例子。注意，$x(t,\zeta_n)$ 就是与试验结果 ζ_n 对应的所有时间$t\in\mathcal{T}$上的完整的波形。

例 6-1 考虑一个随机过程，其在 $t\in\mathcal{T}=\mathcal{R}$上的所有现实都是幅值不同的余弦波。对一个特定的 $\zeta\in\mathcal{S}$，过程的现实为：

$$x(t,\zeta)=\zeta^a\cos(2\pi f_c t) \tag{6-1}$$

其中 ζ^a 是与现实 ζ 对应的幅值。例如，ζ^a 的值在 $[0,1]$ 间均匀分布。这个随机过程具有某些结构(随机性不是很强)，因为在波形的集合中只有余弦波的幅值是随机的。对这种有结构的情况，我们可将样本空间用更简单的形式来描述；用 $S=[0,1]$ 表示随机过程所有可能的现实，试验结果由式(6-1)给出。实际上，这个随机过程在本质上与 $[0,1]$ 区间上均匀分布的随机变量等价。假设所有的现实还有不同的相移：

$$x(t,\zeta)=\zeta^a\cos(2\pi f_c t+\zeta^p) \tag{6-2}$$

其中 $\{\zeta^a,\zeta^p\}$ 为现实 ζ 的幅度和相位。对这样的现实，样本空间应该用 $S=[0,1]\times[-\pi,\pi]$ 描述，其中 $[0,1]$ 和 $[-\pi,\pi]$ 分别对应可能的幅值和相位。可以理解为一个试验结果 $\zeta\in\mathcal{S}$对应一个特定幅度和相位的波形。这个随机过程本质上等价于一个具有两个独立元素的随机向量，这两个元素分别在 $[0,1]$ 和 $[-\pi,\pi]$ 上均匀分布。◀

通常，每个 $\zeta\in\mathcal{S}$对应于一个特定的现实(波形)，如图 6-2 所示。但是，如果波形具有例 6-1 所示的某种结构，我们可以用波形的随机参数(比如幅度或相位)来更简单的描述$\mathcal{S}$，以区分所有的现实。这种简单的情况有时与随机变量或随机向量等价。这章的后面将会学习具有更复杂结构的随机过程。

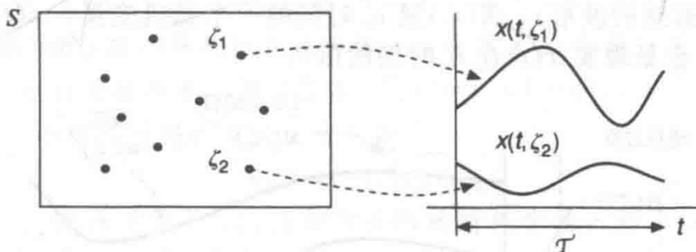


图 6-2 在整个时间间隔$\mathcal{T}$的随机过程的现实可被看作$\mathcal{S}$的输出

随机过程也可看做是试验(随机变量)在时间上的连续输出。对应每一个时刻(连续或离散)存在一个基础概率空间 $\{\Omega,\mathcal{F},P\}$，它在整个$\mathcal{T}$集上可能相同，也可能不同(σ-域$\mathcal{F}$和概率赋值$\mathcal{P}$与上文中用样本空间$\mathcal{S}$来描述的随机过程不同)。例如，重复投掷硬币生成一个随机序列，此序列的基础样本空间是固定的。对每个 $t\in\mathcal{T}$，其试验结果可由当时的概率分布来"描绘"。因此，随着时间的推移，随机过程的现实由各个独立随机试验的结果构成。用这种方法来生成一个离散时间的序列和具有两个试验结果的 Ω 的例子见图 6-3。观察到 $x(3,\omega_2)$ 是试验结果 ω_2 在 $t=3$ 时刻的值。注意到图 6-2 中样本空间$\mathcal{S}$中的元素对应的是随机过程的现实是非常重要的，而图 6-3 中 Ω 的元素是试验在某个特定时刻的输出：$X(3,\omega)$ 是时刻 3 对应的随机变量，其当前的输出为 $x(3,\omega_2)$。

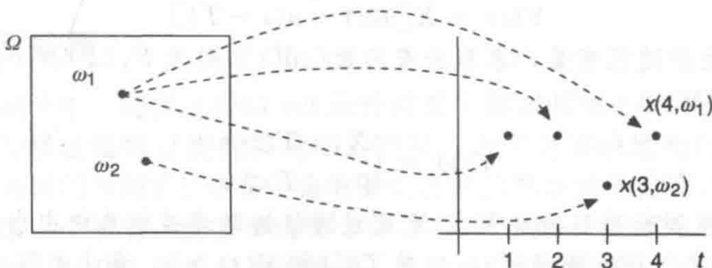


图 6-3 一个随机序列在不同时刻的值可被看作是随机变量的输出

第二个基于 Ω 的描述代表了随机过程的现实实际上是如何产生的，例如上文所提到的重复投掷硬币产生了一个正面和反面的序列。在$\mathcal{T}$上的这些试验结果是伯努利序列的一个现实(稍后再详细讨论)。如果重新开始试验，观察连续投掷硬币的结果，则几乎可以确定会得到过程的一个不同的现实。尽管所有的现实确实是这样产生的，我们仍可以用随机过程的第一个描述，即每个正面和反面序列与随机过程的基础样本空间$\mathcal{S}$的一个独特的元素 ζ 对应。

看另一个基于一天中温度变化的例子。同样，我们可以用这个过程的现实的任何一个描述。在每个时刻，可以认为温度就是某随机变量的输出(可能是一个截顶的高斯分布)。或者，假设温度的所有可能的现实都可以用连续的样本空间$\mathcal{S}$来描述，那一天产生的一个现实就是一个特定的 ζ。与投掷硬币试验不同，这个温度变化的过程的产生不那么明显，它依赖于许多物理因素，而这些物理因素无法都考虑在内。不管怎样，两种描述都是随机过程的有用的表述，我们可用概率方法来检验。

前面的例子说明了随机过程的一个重要并将引起特别关注的特性：不同时刻的输出是否相关。显然对于投掷硬币试验输出是独立的，即，后面的输出不依赖于前面投掷硬币的结果。有些类型的随机过程情况可能不同，它们是高度相关的。温度的例子就是一个输出相关的“试验”。我们可以预测将来某一时刻的温度会与最近时刻观测到的温度非常接近。相关性可能是随机过程诸多应用中最重要的一个特性，第 8 章将进一步讨论，稍后应用的章节也会用到。

图 6-4 给出了随机过程符号的总结。尽管对 ζ 的依赖已在图中清晰显示(以及在例 6-2 的图中)，但在随后的讨论中仍会省略它使符号简化。因此，$X(t)$代表随机过程，$x(t)$是$\mathcal{T}$上单个的现实(观察到的波形)，$X(t_0)$是 t_0 时刻的一个随机变量，$x(t_0)$是随机变量在 t_0 时刻的单个输出(它也是现实 $x(t)$在 t_0 时刻的值)。

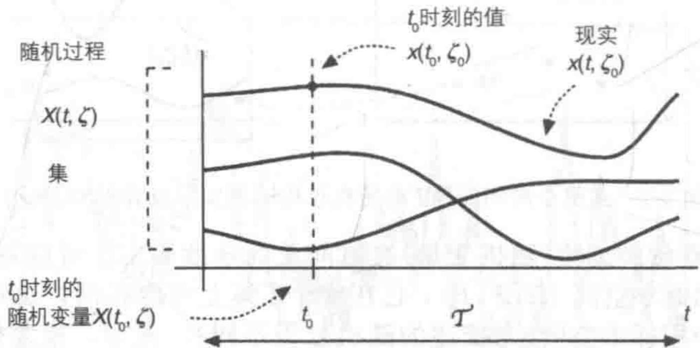


图 6-4　随机过程的符号总结。(ⅰ) $X(t,\zeta)$：包括 $\zeta\in\mathcal{S}$现实集的随机过程。(ⅱ) $x(t,\zeta_0)$：$X(t,\zeta)$ 对应于 $\zeta_0\in\mathcal{S}$的现实(关于时间的确定函数)；(ⅲ) $X(t_0,\zeta)$：在 t_0 时刻所有现实的输出对应的随机变量；(ⅳ) $x(t_0,\zeta_0)$：$\zeta_0\in\mathcal{S}$的现实在 t_0 时刻的值，与随机变量在 t_0 时刻的一个输出对应

例 6-2 在随机过程通用的概率描述之前，考虑另一个相对简单的随机过程，定义如下：

$$Y(t)=X[u(t)-u(t-T)] \tag{6-3}$$

其中 X 和 T 是独立的随机变量，累积分布函数(cdf)分别为 $F_X(x)$和 $F_T(t)$。由于阶跃函数 $u(t)$的作用，得到：

$$Y(t)=\begin{cases}X, & T>t\\ 0, & T\leqslant t\end{cases} \tag{6-4}$$

其中 $t\in\mathcal{T}=\mathcal{R}^+$。这类过程可能在 Y 与 X 通过通信链路连接的系统中出现，T 是链路断开的时刻。因此，在某个特定的时刻 t，如果 $T\leqslant t$ 则 $Y(t)$为 0，意味着链路在 t 之前(或正好在 t 时刻)断开。图 6-5 给出了现实的示例，从中可以看到随机性的两个影响：(ⅰ)X 的输

出决定$Y(t)$的非零幅值(ⅱ)T的输出决定$Y(t)$何时减至零。通过对T设定条件导出其cdf(时间的函数)来描述这个过程的特征，如下所示：

$$P(Y(t)\leqslant y)=P(Y(t)\leqslant y|T\leqslant t)P(T\leqslant t)+P(Y(t)\leqslant y|T>t)P(T>t) \quad (6\text{-}5)$$

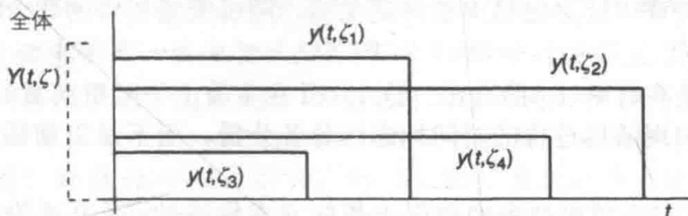


图6-5 例6-2中随机过程的现实示例。每个ζ_n表示一个特定的X和T的输出对

由以上的描述：

$$P(Y(t)\leqslant y|T\leqslant t)=u(y) \quad (6\text{-}6)$$

因为在此条件下$Y(t)=0$，因此其依概率1小于任意正数y。对于其他条件概率：

$$P(Y(t)\leqslant y|T>t)=P(X\leqslant y)=F_X(y) \quad (6\text{-}7)$$

因为链路未断开，$Y(t)=X$。结合以上结果可给出下面的时变cdf：

$$\begin{aligned}F_{Y(t)}(y)&=u(y)F_T(t)+F_X(y)[1-F_T(t)]\\&=F_T(t)[u(y)-F_X(y)]+F_X(y)\end{aligned} \quad (6\text{-}8)$$

其时间依赖性(使$Y(t)$减至零)源于式(6-8)中的第一项。$Y(t)$是一个非平稳过程的例子(稍后定义)，因为其分布随时间而改变。对应的概率密度函数(pdf)为：

$$f_{Y(t)}(y)=F_T(t)\delta(y)+f_X(y)[1-F_T(t)] \quad (6\text{-}9)$$

其中$f_X(y)$是X的pdf，$\delta(y)$是狄拉克δ函数，证明了$Y(t)$在每个有限时间t是一个混合随机变量。它同时也为混合概率，加权值为$\{F_T(t),1-F_T(t)\}$，显然和为1。◀

例6-3 考虑一个随机过程，其现实定义如下：

$$X(t)=A\exp(-\lambda t) \quad (6\text{-}10)$$

其中$t\in\mathcal{R}^+$，$\lambda>0$。设A为在$[0,1]$上均匀分布的随机变量，则$X(t)$的现实集为初始值不同的衰减指数。它具有以下原点矩：

$$\mathcal{E}[X^n(t)]=\mathcal{E}[A^n]\exp(-\lambda nt)=\exp(-\lambda nt)/(n+1) \quad (6\text{-}11)$$

t时刻的cdf为：

$$\begin{aligned}P(X(t)\leqslant x)&=P(A\exp(-\lambda t)\leqslant x)=P(A\leqslant x\exp(\lambda t))\\&=x\exp(\lambda t)I_{[0,1]}(x\exp(\lambda t))+I_{(1,\infty)}(x\exp(\lambda t))\\&=x\exp(\lambda t)I_{[0,\exp(-\lambda t)]}(x)+I_{(\exp(-\lambda t),\infty)}(x)\end{aligned} \quad (6\text{-}12)$$

代入均匀分布的随机变量的cdf得到最后的表达式。将其对x微分得到pdf：

$$f_{X(t)}(x)=\exp(\lambda t)I_{[0,\exp(-\lambda t)]}(x) \quad (6\text{-}13)$$

我们看到其在$[0,\exp(-\lambda t)]$上是均匀分布的。(注意到指示函数只是用来明确地显示定义域，不将它们对x微分)。随着t的增长，由指示函数的下标所给定的pdf的输出范围逐渐向零压缩。◀

6.3 一致性及扩展

在最后两个例子中，随机过程的cdf是针对某个特定时刻t推导得到的。在这些例题中，并不总是直接描述随机过程的各现实的特征。为了更准确地给出随机过程的概率描述，有必要定义在时间间隔$\mathcal{T}$上所有子集的随机变量的联合概率分布。我们可以将随机过程看做随机向量的一个扩展，向量中的每个分量由$\mathcal{T}$中的一个元素来标记，可为无限个元素，可数或不可数。从这个随机向量的描述，就能确定联合分布，从而确定随机变量的联合特性，比如联合矩、独立性、相关性、正交性等。

考虑离散时刻的子集 $\{t_1,\cdots,t_N\}\in\mathcal{T}$，不要求升序排列，而且可与$\mathcal{T}$中初始元素在时间上有差异。尽管通常认为时间是向前移动且没有间隔(当时刻是均匀等分的)，但我们可以自由选择$\mathcal{T}$的任意子集，其元素可为任意排序，因为我们考虑的是一个随机变量的集合。正如我们在第 4 章中针对随机向量所定义的，随机变量集合的联合 cdf 如下所示：

$$F_{X(t_1),\cdots,X(t_N)}(x_1,\cdots,x_N)=P(X(t_1)\leqslant x_1,\cdots,X(t_N)\leqslant x_N) \tag{6-14}$$

其中小写字母$\{x_k\}$是在时刻$\{t_k\}$的输出。这个 cdf 在本质上与随机向量的 cdf 具有相同的形式，但是在这里我们用随机过程的时间标志区分各分量，而不是以前随机向量元素用到的下标。

随机向量的 cdf 与由随机过程的有限子集的元素构造的 cdf 几乎没有区别。为了使一个随机过程的概率模型有意义，$\mathcal{T}$的任意子集的 cdf 必须是一致的。通常来说：(ⅰ)用$\mathcal{T}$的子集进行索引的元素的任意排序的概率测度必须给出明确的定义；(ⅱ)应该非常便捷的由一个更大子集的 cdf 得到这个随机变量的子集的 cdf。为进一步明确该问题，下面给出随机变量子集的概率测度的一致性定义。

定义(概率测度的一致性)　如果对随机变量的有限维分布，以下两个条件成立，则概率测度 P 是一致的：

- 对子集 $\{t_1,\cdots,t_N\}\in\mathcal{T}$ 和随机变量的事件 $\{E_1,\cdots,E_N\}$，设 $\{\pi(1),\cdots,\pi(N)\}$ 为下标的排序，则

  $$P(X(t_{\pi(1)})\in E_{\pi(1)},\cdots,X(t_{\pi(N)})\in E_{\pi(N)})=P(X(t_1)\in E_1,\cdots,X(t_N)\in E_N) \tag{6-15}$$

  如果概率测度中的随机变量和事件是可排序的，则可以得到与随机变量和相关事件初始排序相同的概率。这个条件确保了随机变量的任意有限维子集的概率不依赖于某一特殊的排序：排序不会改变 $P(\cdot)$中变量 $\{E_1,\cdots,E_N\}$ 所给定的全部合成事件。
- 对更大子集 $\{t_1,\cdots,t_N,t_{N+1}\cdots,t_{N+M}\}\in\mathcal{T}$

  $$\begin{aligned}P(X(t_1)\in E_1,\cdots,X(t_N)\in E_N)=&P(X(t_1)\in E_1,\cdots,X(t_N)\in E_N\\&X(t_{N+1})\in\Omega_{N+1},\cdots,X(t_{N+M})\in\Omega_{N+M})\end{aligned} \tag{6-16}$$

  其中 Ω_k 指在 t_k 时刻随机变量的样本空间。由于 $m=1,\cdots,M,X(t_{N+m})\in\Omega_m$ 是确定的事件，则式(6-16)中两个参数的合成事件实际上是相同的，因此概率是不变的。这个条件确保了由一个高阶联合概率导出事件子集的概率的方法的一致性，而且可由联合概率求出边缘概率。

这两个关于概率测度一致性的条件直观上非常吸引人。我们现在给出柯尔莫哥洛夫扩展定理(不做证明)，它指出了与一组有限维分布相关的一个随机过程的存在，也被称为柯尔莫哥洛夫存在性定理。

定理 6-1(柯尔莫哥洛夫扩展)　设概率测度 $P(X(t_1),\cdots,X(t_N))$ 对任意采样时间集合 $\{t_1,\cdots,t_N\}$ 一致，其中 $t_n\in\mathcal{T}$，$N\in\mathcal{N}$可为任意尺寸，则存在一个随机过程 $X(t)$，其中 $t\in\mathcal{T}$，可以产生任意有限维分布。

这个定理将随机变量的有限集合 $\{X(t_1),\cdots,X(t_N)\}$ 扩展到随机过程 $X(t)$，通过(ⅰ)定义时间$\mathcal{T}$的一个集合，使 $t\in\mathcal{T}$；(ⅱ)确保用$\mathcal{T}$的任何有限子集索引的所有随机变量是一致的。$\mathcal{T}$中元素的集合可以是有限的，可数无穷，或不可数(连续的)。该定理与随机过程的第二种表述相关，即从由时间索引的随机变量连续地选择样本得到随机过程。如果该定理的条件满足，则随机过程存在，而且可由一个基础概率空间 $\{S,\mathcal{F},P\}$ 来描述，以随机过程的现实作为其结果。同一个随机过程的两种不同描述方法的关系如下。如果随机变量 $X(t_n)$的样本空间为 $\Omega_n(n\in\mathcal{N})$，则随机过程的样本空间为$\mathcal{S}=\Omega_1\times\Omega_2\times\cdots$。现实 ζ 是随机变量的结果在样本空间$\mathcal{S}$上扩展出的一个特定波形。如果 $\Omega_n=\Omega$ 对所有 n 是时不变的(称为平稳)，则样本空间“简化”为$\mathcal{S}=\Omega\times\Omega\times\cdots=\Omega^{|T|}$，在每个时刻的可能结果是相同的，

$|\mathcal{T}|$是$\mathcal{T}$的势。由之前的定义可知概率测度一定是一致的。通常用各离散时刻的概率测度来表示，除非随机变量是独立的。概率测度包含了任何可能存在的跨越各时刻的依赖关系，这显然不能从一个单一的时刻推算出来。柯尔莫哥洛夫扩展定理将离散时间的描述扩展为连续时间，并允许不同时刻的随机变量的概率空间变化，只要概率测度是一致的。

例 6-4 考虑连续掷硬币，其中$\Omega=\{H, T\}$，为简单起见假定$\mathcal{T}=\{t_1,t_2\}$。随机序列的样本空间为$\mathcal{S}=\{HH,HT,TH,TT\}$，其中时间索引简化为排序的试验结果。或者，可以用掷硬币的样本空间写为$\mathcal{S}=\Omega\times\Omega=\Omega^2$。假设用抛骰子代之抛硬币，在第一时刻，$\Omega_1=\{H, T\}$，而第二时刻$\Omega_2=\{1, 2, 3, 4, 5, 6\}$。因此，不可能将$\mathcal{S}$写为$\Omega^{|\mathcal{T}|}$，因为$\Omega$变化了。这是一个非平稳随机过程，试验结果的分布随时间而改变。我们代之以$\mathcal{S}=\Omega_1\times\Omega_2=\{\{H,1\},\{T,1\},\cdots,\{H,6\},\{T,6\}\}$，再次将时间索引简化为有序的试验结果。◀

6.4 随机过程的类型

接下来，对随机过程的4种基本类型进行总结，这取决于$\mathcal{T}$和随机变量是连续的还是离散的。为了清楚起见，在必要的时候，我们用以下符号区分连续时间过程和离散时间序列：

随机序列：$X[k]$，$k\in\mathcal{T}\subseteq\mathcal{Z}$ (6-17)

随机过程：$X(t)$，$t\in\mathcal{T}\subseteq\mathcal{R}$ (6-18)

其中方括号表示序列(时间序列)，它通常用在对离散时间信号的描述中(确定的或随机的)。这四种类型的随机过程的现实的简单例子见图6-6，其中在离散时间情况下，假定时间为均匀间隔(上面用$\mathcal{Z}$时也是同样的)。

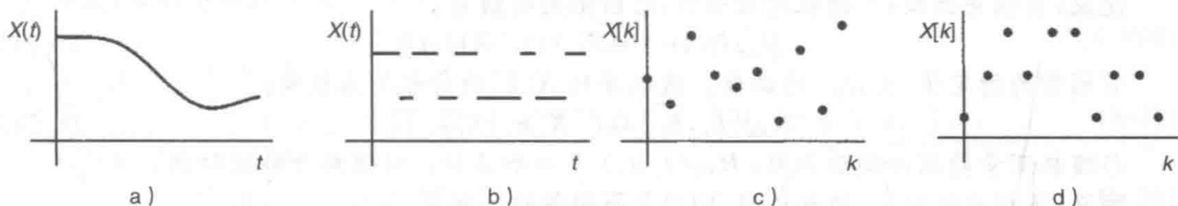


图6-6 随机过程的现实示例。a) 连续时间，连续幅度；b) 连续时间，离散幅度(2个值)；c) 离散时间，连续幅度；d) 离散时间，离散幅度(三个值)

除了图中的四种情况外，基础概率空间有可能是混合的，随机变量由连续和离散部分构成。例6-2中给出了一个简单的混合过程，其中$Y(t)$在$y=0$处具有非零的概率质量。随机过程的其他问题和特性在之前的描述中已比较清楚了。例如，随机变量可能随时间变化(例如从高斯变为指数)，或对所有的$\mathcal{T}$有相同的随机变量存在，但它的一个或多个参数可以改变(如随时间改变均值的高斯变量)。本章中，我们将涉及随机过程的重要例子，讨论它们作为相关模型时的性质及某些应用，并研究随机变量在不同时刻的相关性。

我们从随机过程是由时间索引的随机变量的集合而引出的某些通用的属性或特征开始。

6.5 平稳性

随机过程有几种不同类型的平稳。通常平稳性是指由时间索引的随机变量的概率模型随时间变化为“常数”的程度。

定义(一阶平稳) 如果随机过程$X(t)$的pdf与时间无关，则为**一阶平稳**，即

$$f_{X(t)}(x)=f_X(x),\quad 对所有\ t\in\mathcal{T} \tag{6-19}$$

因此，所有的矩(原点和中心矩)、累积量等都是恒定的。对于这种类型的平稳，我们

不去讨论在两个或更多不同时刻的矩，比如随机量的相关性和协方差等。

定义（N 阶平稳） 随机过程 $X(t)$ 是 N **阶平稳的**，如果过程在 N 个时刻的联合概率分布不随 $\{t_1,\cdots,t_N\}$ 的任意时移 τ 而改变，即

$$f_{X(t_1),\cdots,X(t_N)}(x_1,\cdots,x_N)=f_{X(t_1-\tau),\cdots,X(t_N-\tau)}(x_1,\cdots,x_N) \tag{6-20}$$

这种类型的平稳性明显强于一阶平稳；它意味着时移不会带来不同时刻矩（最多到 $N-1$ 时间差）的改变。我们可以将 N 阶平稳看做一个“滑动窗口”（宽度为 N），滑窗内的联合 pdf 不会改变。

定义（严格平稳） 随机过程 $X(t)$ 是**严格平稳**的，只要过程对所有的 N 满足当 $N\to\infty$ 时 N 阶平稳。

这是平稳的最强形态：随机过程的所有统计特性都不随时间差和时移而改变。严格平稳过程意味着式(6-20)，则也意味着式(6-19)。对于所有三种情况，边缘分布是相同的：分布的类型（高斯分布，泊松分布等）是相同的，pdf 的参数不会改变。一般情况下，式(6-19)并不意味着式(6-20)。一个重要的例外是当随机过程是独立时，所有的联合分布可表示为边缘分布的乘积。这对应于一个独立同分布（iid）的随机过程，将在 6.6 节中讨论。

在许多应用中，三个平稳的定义通常要强于表征随机过程所需的特性。严格平稳随机过程的需求太高以至于对绝大多数实际过程是不必要的。N 阶平稳随机过程的要求也非常高，因为要求 N 个时刻的所有矩都为常数，这在实际中是不现实的。对于许多问题，我们通常只对一阶矩和二阶矩（尤其是相关性）以及它们是否为时不变感兴趣。稍后详细讨论相关性及其频域的对应项（通过傅里叶变换），即功率谱密度（PSD）。

回顾第 5 章随机变量的相关性的定义，并将其应用到随机过程中。

定义（自相关函数） 随机过程 $X(t)$ 的**自相关函数**为：

$$R_{XX}(t_1,t_2)\triangleq\mathcal{E}[X(t_1)X(t_2)] \tag{6-21}$$

它通常为特定值 $\{t_1,t_2\}$ 的函数。随机序列 $X[k]$ 的自相关函数为：

$$R_{XX}[k_1,k_2]\triangleq\mathcal{E}[X[k_1]X[k_2]] \tag{6-22}$$

与随机变量的相关函数不同，$R_{XX}(t_1,t_2)$ 是一个函数：它依赖于特定时刻。

定义（不相关过程） 随机过程 $X(t)$ 是**不相关**的，如果

$$R_{XX}[t_1,t_2]=\begin{cases}\mathcal{E}[X^2(t_1)], & t_1=t_2\\ \mathcal{E}[X(t_1)]\,\mathcal{E}[X(t_2)], & t_1\neq t_2\end{cases} \tag{6-23}$$

同样对随机序列 $X[k]$：

$$R_{XX}[k_1,k_2]=\begin{cases}\mathcal{E}[X^2(k_1)], & k_1=k_2\\ \mathcal{E}[X(k_1)]\,\mathcal{E}[X(k_2)], & k_1\neq k_2\end{cases} \tag{6-24}$$

如果过程是零均值，则自相关函数也为零，除非变量相等，则对应的是特定时刻的方差值。

定义（自协方差函数） 随机过程 $X(t)$ 的**自协方差函数**为：

$$C_{XX}[t_1,t_2]\triangleq\mathcal{E}[(X[t_1]-\mathcal{E}[X[t_1]])(X[t_2]-\mathcal{E}[X[t_2]])] \tag{6-25}$$

随机序列 $X[k]$ 的自协方差函数：

$$C_{XX}[k_1,k_2]\triangleq\mathcal{E}[(X[k_1]-\mathcal{E}[X[k_1]])(X[k_2]-\mathcal{E}[X[k_2]])] \tag{6-26}$$

如果过程是不相关的，则自协方差函数为零，除非变量相等，则此时对应的是方差值（即使对非零均值）。接下来讨论的平稳类型通常假设在许多工程应用中对随机信号进行操作的实际系统中成立。

定义（广义平稳） 随机过程 $X(t)$ 是**广义平稳**的，如果均值为常数，自协方差函数只与时间差相关，而不依赖于特定时刻：

对所有 $t\in\boldsymbol{T}$，满足 $$\mathcal{E}[X(t)]=\mu_X \tag{6-27}$$

对所有 $t_1,\ t_2\in\boldsymbol{T}$，满足 $$C_{XX}(t_1,\ t_2)=C_{XX}(t_2-t_1) \tag{6-28}$$

时间差 t_2-t_1 也称为时间间隔，当过程为广义平稳时，通常用 $\tau \triangleq t_2-t_1$ 来代替时间差，即

$$C_{XX}(t_2-t_1)=C_{XX}(\tau) \tag{6-29}$$

对随机序列，有

$$C_{XX}[k_2-k_1]=C_{XX}[m] \tag{6-30}$$

其中 $m \triangleq k_2-k_1$。相应的自相关函数为 $R_{XX}(\tau)$ 和 $R_{XX}[m]$。

对不相关过程，$C_{XX}(\tau)=\sigma_X^2\delta(\tau)$，$C_{XX}[m]=\sigma_X^2\delta[m]$。这将在第 8 章进行深入讨论，还将给出白噪声的定义。

例 6-5 例 6-3 中随机过程的自相关函数为：

$$R_{XX}(t,t+\tau)=\mathcal{E}[A^2]\exp(-\lambda(2t+\tau))=(1/3)\exp(-\lambda(2t+\tau)) \tag{6-31}$$

证明了 $X(t)$ 不是广义平稳过程。均值也随时间改变：$\mu_X(t)=(1/2)\exp(-\lambda t)$。 ◀

广义平稳过程将在第 8 章广泛讨论，通常在应用章节中会假设各种信号为广义平稳。

从定义中的对称性可推出自相关函数是偶函数：$R_{XX}(\tau)=R_{XX}(-\tau)$，显然对于时间间隔 0，它是非负的：$R_{XX}(0)=\mathcal{E}[X^2(t)]\geqslant 0$。接下来的定理描述了自相关函数的其他特性。

定理 6-2 自相关函数在时间间隔为 0 时最大：

$$|R_{XX}(\tau)|\leqslant R_{XX}(0) \tag{6-32}$$

证明： 由附录 F 中柯西-施瓦兹不等式可得：

$$|R_{XX}(\tau)|=|\mathcal{E}[X(t)X(t+\tau)]|\leqslant\sqrt{\mathcal{E}[X^2(t)]\,\mathcal{E}(X^2(t+\tau)]} \tag{6-33}$$

由于过程是广义平稳的，$\mathcal{E}[X^2(t)]=\mathcal{E}[X^2(t+\tau)]=R_{XX}(0)$，证明完毕。

之前关于相关和协方差的定义可扩展应用于两个随机过程。

定义（互相关和协方差） 随机过程 $X(t)$ 和 $Y(t)$ 的互相关函数为：

$$R_{XY}(t_1,t_2)\triangleq\mathcal{E}[X(t_1)Y(t_2)] \tag{6-34}$$

互协方差函数为：

$$C_{XY}(t_1,t_2)\triangleq\mathcal{E}[(X(t_1)-\mathcal{E}[X(t_1)])(Y(t_2)-\mathcal{E}[Y(t_2)]] \tag{6-35}$$

对随机序列 $X[k]$ 和 $Y[k]$，

$$R_{XY}[k_1,k_2]\triangleq\mathcal{E}[X[k_1]Y[k_2]] \tag{6-36}$$

$$C_{XY}[k_1,k_2]\triangleq\mathcal{E}[(X(k_1]-\mathcal{E}[X[k_1]])(Y[k_2]-\mathcal{E}[Y[k_2]])] \tag{6-37}$$

当过程为广义平稳时，这些表达式变为 $R_{XY}(\tau)$，$C_{XY}(\tau)$，$R_{XY}[m]$ 和 $C_{XY}[m]$，其中 $\tau\triangleq t_2-t_1$，$m\triangleq k_2-k_1$

注意到广义平稳随机过程的自相关函数的对称性在这并不适用：例如 $R_{XY}(\tau)\neq R_{XY}(-\tau)$，反而有 $R_{XY}(\tau)=R_{YX}(-\tau)$。

为保证完整性，我们还给出了周期平稳的定义；它只在第 7 章中偶尔用到。

定义（周期平稳） 随机过程 $X(t)$ 是**广义周期平稳**的，如果其均值为常数，自协方差函数是以 T_o 为周期的周期函数，即对任意 τ 有：

$$C_{XX}(\tau)=C_{XX}(\tau+T_o) \tag{6-38}$$

6.6 独立同分布

因为一个随机过程被看作是由时间索引的随机变量的集合，我们可以将独立随机变量的概念扩展到随机过程。考虑一个独立随机序列的以下示例。

例 6-6 在 $k\in\mathcal{T}=\mathcal{Z}^+$ 的每个时刻，随机序列 $X[k]$ 的取值 $x[k]$ 由一个具有固定参数 p 的伯努利随机变量决定。这与以概率 $P(H)=p$ 进行的投掷硬币等价。在每次投掷硬币时观察试验结果，并将结果“加入”到之前的试验结果中构造一个现实（一个离散时间函数）。随机变量的 pdf 对 k 保持不变，这意味着随机序列是一阶平稳的。由于每次投掷是独立的，随机序列也是严格平稳的。假设硬币被投掷 N 次，，则随机序列 $X[k]$ 有 2^N 种可能的现

实。图 6-7 显示了一个例子，其中 $H\equiv 1$，$T\equiv 0$。这 2^N 种可能的现实集就是随机序列的总体：每个现实是基本概率空间的一个结果。对这种离散的、有限值的例子，可以非常容易的将随机序列的样本空间用伯努利随机变量的样本空间 $\Omega=\{H,T\}$ 表示为如下的笛卡儿积：$\mathcal{S}=\Omega\times\cdots\times\Omega=\Omega^N$。它的元素 $\{\zeta\}$ 有 2^N 种排列：$\{H,\cdots,H,H\},\{H,\cdots,H,T\},\cdots,\{T,\cdots,T,T\}$。由于序列是按时间索引的（意味着一个方向），我们必须考虑所有的排序，而不是简单的组合；现实 $\{H,\cdots,H,H,T\}$ 不同于 $\{H,\cdots,H,T,H\}$，即便其中都只有一个 T。由于随机变量是独立的、平稳的（p 不会改变），每个 $\zeta\in\mathcal{S}$（m 个 H）的概率为 $p^m q^{N-m}$，其中 $q=1-p$。

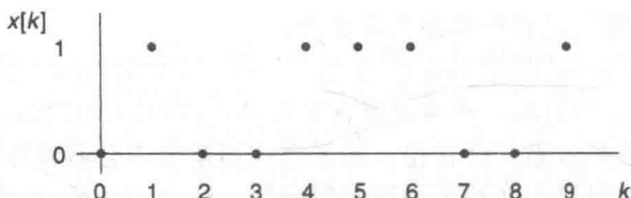


图 6-7　一个 $\mathcal{T}=\{0,\cdots,9\}$ 的伯努利随机序列的现实 $x[k]$ 的例子

很显然，这个例子中的概率测度是一致的。$p^m q^{N-m}$ 不依赖于随机变量的结果的顺序：由 m 个 H 和 $N-m$ 个 T 组成的任何排列具有相同的概率。考虑 N 个随机变量的一个大小为 M 的子集；这个子集的概率的计算方法如下。如果该子集包含 m 个 H 和 $M-m$ 个 T，则

$$P(\underbrace{H,H,\cdots,H}_{m项},\underbrace{T,\cdots,T}_{M-m项})=P(\underbrace{H,H,\cdots,H}_{m项},\underbrace{T,\cdots,T}_{M-m项},\underbrace{\Omega,\cdots,\Omega}_{N-M项})$$
$$=p^m q^{M-m}1^{N-M}=p^m q^{M-m} \tag{6-39}$$

其中我们运用了独立性和 $P(\Omega)=1$。更小子集的概率可通过包含 $N-M$ 个确定事件 Ω 的更大的子集的概率计算得到。◀

由这个例子，我们发现一致性定义第二部分的内容中关于通用分布可陈述如下：考虑随机过程 $X(t)$ 在 N 个时刻的联合 cdf 为 $F_{X(t_1),\cdots,X(t_N)}(x_1,\cdots,x_N)$。则对应于 $M<N$ 时刻的子集的联合 cdf 可由下式得到：

$$F_{X(t_1),\cdots,X(t_M)}(x_1,\cdots,x_M)=F_{X(t_1),\cdots,X(t_M),X(t_{M+1}),\cdots,X(t_N)}(x_1,\cdots,x_M,\infty,\cdots,\infty) \tag{6-40}$$

其中，不失一般性，我们假设 $t_M<t_N$，相应的联合 pdf 为：

$$f_{X(t_1),\cdots,X(t_M)}(x_1,\cdots,x_M)=\underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{N-M积分}f_{X(t_1),\cdots,X(t_{M+1}),\cdots,X(t_N)}(x_1,\cdots,x_{M+1},\cdots,x_N)\mathrm{d}x_{M+1}\cdots\mathrm{d}x_N \tag{6-41}$$

第 4 章涉及随机向量时出现过这种类型的表达式。它们的区别是随机过程的这些性能对 $\mathcal{T}$ 中的所有子集必须成立，以便具有一致性。当然，这些特性不需要进行验证，只需对于特定的应用指定随机过程的类型，加上概率测度一致的约束。

我们已知伯努利随机序列是独立同分布（iid）的，这也许是最基本的一种随机过程。

定义（IID 随机过程）　如果在所有 $t\in\mathcal{T}$ 时刻的随机变量相互独立，并且具有相同的分布 $F_X(t)(x)=F_X(x)$，则随机过程 $X(t)$ 是 iid。

如例 6-5 中伯努利随机序列所示，一个长度为 N 的 iid 随机序列样本空间的形式为 $\mathcal{S}=\Omega^N$。

例 6-7　连续投掷骰子 $k\in\mathcal{Z}^+$ 次，得到一个 iid 随机序列。除了随机变量满足具有 6 个输出 $\Omega=\{1,2,3,4,5,6\}$ 的离散的均匀分布，其他与伯努利随机序列类似。◀

例 6-8　大多数信号处理软件（如 MATLAB）用的随机数生成器是一个构造 iid 随机序列算法的例子（虽然在技术上它是一个伪随机数（PRN）产生器，因为最终数字会重复，尽

管仿真中通常不会达到这点)。一个理想的均匀随机数生成器会生成 $[0,1]$ 间 iid 的数字序列。用第 4 章中描述的技术可以使这个序列的分布变为一个不同分布。◀

例 6-9 高斯随机序列 $X[k]$ 是 iid 的，只要满足任意由 $k(k\in\mathcal{T}=\mathcal{Z}^+)$ 索引的随机变量的有限集合是相互独立的，并且都有固定参数为 $\{\mu,\sigma\}$ 的相同的高斯分布。一般高斯随机过程需要随机变量集是联合高斯的；对 iid 的情况，则只需要每个随机变量是高斯的，因为联合 pdf 是边缘 pdf 的乘积，而边缘概率密度是相同的。◀

虽然 iid 随机过程在实际中得到了广泛应用(如数字通信中)，但事实上若要求在不同时间的样本是独立的则过于严格。在此过程中没有"可预见性"，前面的输出不能提供对未来的输出进行预测的信息。在许多应用中对输出进行某种程度的预测是很重要的，这部分内容我们将在后面的章节中进行讨论。

6.7 独立增量

本节将考虑随机过程的一个重要类型，它可以看作是 iid 过程的扩展，一些大家熟知的随机过程是其中的特例。设 $X(t)$ 是一个随机过程，$t\in\mathcal{T}$。由于 $X(t)$ 在任意时刻是随机变量，固定时刻 t_1 和 t_2 之间的差值(增量) $X(t_1)-X(t_2)$ 也是一个随机变量。利用第 4 章中随机变量变换的方法可得到以下结果：可由 $X(t_1)$ 和 $X(t_2)$ 的 pdf 计算出 $X(t_1)-X(t_2)$ 的 pdf。

定义(独立增量过程) 随机过程 $X(t)$ 具有独立增量，如果对于任意 $t_1<t_2<t_3<t_4$ 有：

$$f_{X(t_2)-X(t_1),X(t_4)-X(t_3)}(x_2-x_1,x_4-x_3)=f_{X(t_2)-X(t_1)}(x_2-x_1)f_{X(t_4)-X(t_3)}(x_4-x_3) \quad (6\text{-}42)$$

这个定义是用于随机过程不相交增量中独立随机变量的一个直接扩展。我们注意到这个增量可以是连续的，各自的时刻是重叠的，比如以下的 t_2 和 t_3：

$$\begin{aligned}&f_{X(t_2)-X(t_1),X(t_3)-X(t_2),X(t_4)-X(t_3)}(x_2-x_1,x_3-x_2,x_4-x_3)\\&=f_{X(t_2)-X(t_1)}(x_2-x_1)f_{X(t_3)-X(t_2)}(x_3-x_2)f_{X(t_4)-X(t_3)}(x_4-x_3)\end{aligned} \quad (6\text{-}43)$$

该定义也适用于下述情况：t_1 与原点(过程的起点)有特定关系，其中 $X(t_0)=x_0=0$ 是固定的(非随机的)。则

$$f_{X(t_1)-X(t_0),X(t_2)-X(t_1)}(x_1-x_0,x_2-x_1)=f_{X(t_1)}(x_1)f_{X(t_2)-X(t_1)}(x_2-x_1) \quad (6\text{-}44)$$

独立增量过程的一些特例会在稍后给出。首先，我们将这类过程扩展到包括平稳性。

定义(ISI 随机过程) 一个独立、平稳增量(isi)过程的增量 $X(t_N-\tau)-X(t_M-\tau)$(对任意 $t_N>t_M$)的分布不会因为任意时移 τ 而改变。

此平稳性只适用于随机过程的增量，而不适用于过程本身。其结果是，相同持续时间的增量是 iid 的。此特性由严格平稳的定义得出，而没有必要考虑增量的所有的联合分布(证明严格平稳性时一般需要)，因为增量是独立的。因为联合矩分成了矩的乘积(由于联合 pdf 为边缘 pdf 的乘积)，一个一阶平稳的独立随机过程也是严格平稳。这同样适用于一个 isi 过程的增量。

接下来，我们考虑对随机序列同样有用的一些分布的表示。

定义(无限可分分布) $F_X(x)$ 是**无限可分分布**的，如果存在 iid 随机变量 $\{X_{n,1},\cdots,X_{n,n}\}$ 使得：

$$Y_n=\sum_{m=1}^{n}X_{n,m},\quad 对每个\ n\in\mathcal{N} \quad (6\text{-}45)$$

具有与 X 相同的分布，则它们是等价的随机变量。

$\{X_{n,m}\}$ 的双下标是用来强调 $\{Y_n\}$ 是由不同的随机变量的和得到的。否则，如果只有一个下标，我们可以写为 $Y_n=Y_{n-1}+X_n$，则 Y_n 和 Y_{n-1} 不具有定义所要求的相同的分布。由于随机变量是 iid 的，它们的和的分布可由相同的 pdf 的 $n-1$ 次卷积得到。如果 X 具有有限定义域，则它不是无限可分的。例中包括了贝塔分布和连续的以及离散的均匀随机变

量。运用特征函数(CF)，卷积对应于相乘：

$$\Phi_X(\omega)=\Phi_{Y_n}(\omega)=\prod_{m=1}^{n}\Phi_{X_{n,m}}(\omega)=\Phi_{X_n}^{n}(\omega) \tag{6-46}$$

由于各 n 时刻的随机变量是iid的，所以得到了上式中最后一个等式。因此，如果对每个 $n\in\mathcal{N}$，X 的CF可以被写为另一个CF的幂，则 X 是无限可分的。表6-1总结了附录A中具有这种特性的几种分布。

表6-1 无限可分分布的例子

随机变量 X\{参数\}	特征函数：$\Phi_X(\omega)$ 或 $\Phi_X(\mathrm{j}\omega)$
柯西分布$\{c,\alpha\}$	$\exp(\mathrm{j}\omega c-\alpha\lvert\omega\rvert)$
伽马分布$\{\lambda,r\}$	$[\lambda/(\lambda-\mathrm{j}\omega)]^r$
几何分布$\{p\}$	$p[1-q\exp(\mathrm{j}\omega)]^{-1}$
高斯分布$\{\mu,\sigma\}$	$\exp(\mathrm{j}\omega\mu-\sigma^2\omega^2/2)$
负二项分布$\{p,N\}$	$p^N/[1-q\exp(\mathrm{j}\omega)]^N$
泊松分布$\{\alpha\}$	$\exp(\alpha(\exp(\mathrm{j}\omega)-1))$
学生氏 t 分布 $t\{r\}$	$K_{r/2}(\sqrt{r}\lvert\omega\rvert)\sqrt{(\sqrt{r}\lvert\omega\rvert)^r}/\Gamma(r/2)2^{r/2-1}$

例6-10 考虑参数为$\{\lambda,r\}$的伽马随机变量，表中给出了其CF：

$$\Phi_X(\omega)=\left(\frac{\lambda}{\lambda-\mathrm{j}\omega}\right)^r=\left[\left(\frac{\lambda}{\lambda-\mathrm{j}\omega}\right)^{r/n}\right]^n \tag{6-47}$$

其中$\{X_{n,1},\cdots,X_{n,n}\}$是参数为$\{\lambda,r/n\}$的iid伽马随机变量。这个例子说明了 n 不同，随机变量集不同，因此 $X_{n,m}$ 需要两个下标。同样，对高斯随机变量：

$$\begin{aligned}\Phi_X(\omega)&=\exp(\mathrm{j}\omega\mu-\sigma^2\omega^2/2)=(\exp([\mathrm{j}\omega\mu-\sigma^2\omega^2/2]/n))^n\\&=[\exp(\mathrm{j}\omega(\mu/n)-(\sigma^2/n)\omega^2/2)]^n\end{aligned} \tag{6-48}$$

其中$\{X_{n,1},\cdots,X_{n,n}\}$是参数为$\{\mu/n,\sigma^2/n\}$的iid高斯随机变量。◀

有些随机变量的无限可分性不容易被直接看出，如学生氏 t 分布。还要注意的是$\{X_{n,1},\cdots,X_{n,n}\}$可能与 X 不具有同类型的分布，对几何随机变量可以证明这点。如果将定义域移到$\mathcal{Z}$(而不是$\mathcal{N}$)(Johnson, Kemp, and Kotz, 2005)，则对数随机变量是无限可分的。

对于一个isi随机过程，任何增量可以分为更小增量的和。特别地，长度 t_2-t_1 的时间间隔可被划分成 N 个连续的子区间，每个长度为$(t_2-t_1)/N$。这些较小的增量也是独立且平稳的，因此一个isi随机过程具有无限可分性。关于随机变量无限可分性的讨论导致了isi随机序列的下列特性。

定理6-3 当且仅当随机序列 $X[k]$ 可以被表示为iid随机变量之和时，它是一个isi序列。

证明： 必要性：设 $X[k]$ 是一个isi序列，$k\in\mathcal{Z}^+$，定义 $Y[k]\triangleq X[k]-X[k-1]$，这必然是一个iid序列。观察可得：

$$X[k]=\sum_{n=1}^{k}(X[n]-X(n-1])=\sum_{n=1}^{k}Y[n] \tag{6-49}$$

其中假设 $X[0]=0$。第一个等式成立是因为所有的项都被抵消了，只留下 $X[k]$。因此 $X[k]$ 被表示为iid随机变量的和。

充分性：如式(6-49)第二个求和项所示，$X[k]$ 为iid随机变量$\{Y[n]\}$的和，则对任意 $k_2>k_1$ 观察可得：

$$X[k_2]-X[k_1]=\sum_{n=1}^{k_2}Y[n]-\sum_{n=1}^{k_1}Y[n]=\sum_{n=k_1+1}^{k_2}Y[n] \tag{6-50}$$

这说明所有的不相交增量是独立的，因为它们不包括$\{Y[n]\}$的任何公共项。任何增量的分布也不会因为时移而改变，因为$\{Y[n]\}$分布相同，证明完毕。

注意到不同尺寸的增量通常具有不同的分布，独立、平稳增量则不必是同分布的。

6.8 鞅

本节中，我们将展示随机序列一个有趣的特性，即随机序列可被分解为两个不同的分量，它在第7章中将会非常有用。为了不失一般性，假设过程从$t=0$时刻开始。

定义(鞅序列) $\mathcal{E}[|X[k]|]<\infty$的随机序列$X[k]$是一个鞅，如果对所有$k\geqslant1$，有：

$$\mathcal{E}[X[k]\,|\,X[0],\cdots,X[k-1]]=X[k-1] \tag{6-51}$$

一定要注意到等号右边是没有求期望值的：这个特殊的条件均值是随机序列最近的随机变量$X[k-1]$。重写这个表达式，观察得到：

$$\mathcal{E}[X[k]-X[k-1]\,|\,X(0),\cdots,X[k-1]]=0 \tag{6-52}$$

其中$X[k-1]$被移入求期望运算中，因为它也是条件作用的部分。这个表达式说明通常鞅序列中从一个时刻到下一个时刻变化为0。

例 6-11 设$Y[k]$是一个伯努利随机序列，其中以概率$p=P(H)$重复投掷一个硬币，定义$X[k]$是k时刻一个赌徒赢钱的净利总和。当硬币出现正面(H)，赌徒赢得固定的金额$c>0$；否则赌徒失去同样的金额c。因此：

$$X[k+1]=\begin{cases}X[k]+c, & Y[k]=H\\ X[k]-c, & Y[k]=T\end{cases} \tag{6-53}$$

以之前所有净盈利为条件，我们得到：

$$\begin{aligned}\mathcal{E}[X[k+1]\,|\,X[0],\cdots,X[k]]&=p(X[k]+c)+q(X[k]-c)\\&=X[k]+(p-q)c\end{aligned} \tag{6-54}$$

其中$q\triangleq1-p$。只有在游戏公平的情况下，即$p=q=1/2$时$X[k]$是一个鞅。◀

对例6-12和例6-13，鞅序列的定义归纳如下。

定义(扩展的鞅序列) 随机序列$Y[k]$，$\mathcal{E}[|Y[k]|]<\infty$，是一个关于序列$X[k]$的鞅，如果对所有$k\geqslant1$，满足：

$$\mathcal{E}[Y[k]\,|\,X[0],\cdots,X[k-1]]=Y[k-1] \tag{6-55}$$

等号右边是$Y[k]$之前的样本，不是条件序列。

例 6-12 参数估计中一个有用的量是似然函数(见第9章)。假设我们对估计一个iid随机序列的pdf的标量参数θ有兴趣。例如，参数可能是一个高斯随机变量的μ或是投掷硬币得到正面的概率p。考虑参数的取值，得到两种假设：H_0：$\theta=\theta_0$和H_1：$\theta=\theta_1$。也有可能有多种假设，但为简单起见，在这个例子中我们只考虑两种假设。似然比定义为：

$$\Lambda(x)=\frac{f_X(x;\theta_0)}{f_X(x;\theta_1)} \tag{6-56}$$

其中参数的值明确的显示在每个pdf的变量中。考虑随机序列的k个时刻$\{X[1],\cdots,X[k]\}$对应的独立随机变量，这k个随机变量的似然比归纳为(Larson和Shubert，1979)

$$\Lambda[k]\triangleq\prod_{m=1}^{k}\frac{f_X(X[m];\theta_0)}{f_X(X[m];\theta_1)} \tag{6-57}$$

它本身也可被看做是一个随机序列。$\Lambda[k]$中的随机性是因为pdf被写为随机变量$\{X[m]\}$的函数(而不是它们的试验结果)。在这种估计问题中，pdf的形式是已知的，只有参数需要估计。对这个例题，θ有两个值都是已知的，我们需要决定的是在观测样本给定条件下θ的哪个值最有可能出现。考虑以下$\Lambda[k]$的条件期望：

$$\mathcal{E}[\Lambda[k]\,|\,X[1],\cdots,X[k-1]]=\mathcal{E}\left[\frac{f_X(X[k];\theta_0)}{f_X(X[k];\theta_1)}\prod_{m=1}^{k-1}\frac{f_X(X[m];\theta_0)}{f_X(X[m];\theta_1)}\,|\,X[1],\cdots,X[k-1]\right] \tag{6-58}$$

将 $\Lambda[k]$ 的值代入上式，并将 k 时刻 pdf 的比值分离出来。由于随机变量是独立的，期望值分解为乘积：

$$\mathcal{E}[\Lambda[k]\,|\,X[1],\cdots,X[k-1]]=\mathcal{E}\left[\prod_{m=1}^{k-1}\frac{f_X(X(m];\theta_0)}{f_X(X[m];\theta_1)}\,|\,X[1],\cdots,X[k-1]\right]\cdot\mathcal{E}\left[\frac{f_X(X[k];\theta_0)}{f_X(X(k];\theta_1)}\right] \tag{6-59}$$

第二个期望值的条件不需要，因为 $X[k]$ 不是条件的部分，而随机变量是独立的。由于第一个期望值是 $\Lambda[k-1]$，则：

$$\mathcal{E}[\Lambda[k]\,|\,X[1],\cdots,X[k-1]]=\Lambda[k-1]\,\mathcal{E}\left[\frac{f_X(X[k];\theta_0)}{f_X(X[k];\theta_1)}\right] \tag{6-60}$$

当且仅当等号右边的期望等于 1 时，可以证明 $\Lambda[k]$ 是关于 $X[k]$ 的一个鞅。这种情况发生在实际参数为 θ_1 时，因此：

$$\mathcal{E}\left[\frac{f_X(X[k];\theta_0)}{f_X(X[k];\theta_1)}\right]=\int_{-\infty}^{\infty}\frac{f_X(x_k;\theta_0)}{f_X(x_k;\theta_1)}f_X(x_k;\theta_1)\,\mathrm{d}x_k=\int_{-\infty}^{\infty}f_X(x_k;\theta_0)\,\mathrm{d}x_k=1 \tag{6-61}$$

如果 $\Lambda[k]$ 是由 θ_0 和 θ_1 互换来定义，则 θ_0 会在以上的结果中用到，如果 $\Lambda[k]$ 的分母依赖于实际参数，则其为鞅。◀

例 6-13（棣莫弗(de Moivre)的鞅） 当例 6-11 中 $p\neq q$ 时，仍有可能通过定义以下序列找到一个鞅：

$$Y[k]=(q/p)^{X[k]} \tag{6-62}$$

其中 $X[k]$ 是初始的伯努利序列，则：

$$\begin{aligned}\mathcal{E}[Y[k+1]\,|\,X[1],\cdots,X[k]]&=\mathcal{E}[(q/p)^{X[k+1]}\,|\,X[1],\cdots,X[k]]\\&=p(q/p)^{X[k]+c}+q(q/p)^{X[k]-c}\end{aligned} \tag{6-63}$$

其中用到了式(6-53)中的条件。将表达式重写为：

$$\begin{aligned}\mathcal{E}[Y[k+1]\,|\,X[1],\cdots,X[k]]&=(q/p)^{X[k]}[p(q/p)^c+q(q/p)^{-c}]\\&=Y[k][p(q/p)^c+q(p/q)^c]\end{aligned} \tag{6-64}$$

我们可以看到 $Y[k]$ 是一个相对于 $X[k]$ 的鞅，只要 $c=1$，即在每个时刻 k，赌徒必须赢或输 1 个单位量。◀

以下定理是鞅序列的另一种扩展，其条件中最近的随机样本不需要是前一时刻的。

定理 6-4 对鞅序列 $X[k]$ 和 $n>k$，对 $1<k\leqslant n$，有

$$\mathcal{E}[X[n]\,|\,X[0],\cdots,X[k-1]]=X[k-1] \tag{6-65}$$

证明：包括一个以 $\{X[0],\cdots,X[n-1]\}$ 为条件的内部望值，则

$$\begin{aligned}\mathcal{E}[X[n]\,|\,X(0],\cdots,X[k-1]]&=\mathcal{E}[\mathcal{E}[X[n]\,|\,X[0],\cdots,X(n-1]]\,|\,X[0],\cdots,X[k-1]]\\&=\mathcal{E}[X[n-1]\,|\,X[0],\cdots,X[k-1]]\end{aligned} \tag{6-66}$$

第二个等式成立是因为假设 $X[k]$ 是一个鞅。将另一个以 $\{X[0],\cdots,X[n-2]\}$ 为条件的内部期望包括进来，则最后一个表达式变为：

$$\begin{aligned}\mathcal{E}[X[n-1]\,|\,X[0],\cdots,X[k-1]]&=\mathcal{E}[\mathcal{E}[X[n-1]\,|\,X[0],\cdots,X[n-2]]\,|\,X[0],\cdots,X[k-1]]\\&=\mathcal{E}[X[n-2]\,|\,X[0],\cdots,X[k-1]]\end{aligned} \tag{6-67}$$

继续对 $X[n-3]$，$X[n-4]$ 等重复这个过程，直到得到 $X[k-1]$，式(6-66)的等号左边变为：

$$\mathcal{E}[X[n]\,|\,X[0],\cdots,X[k-1]]=\mathcal{E}[X(k-1]\,|\,X[0],\cdots,X[k-1]]=X[k-1] \tag{6-68}$$

证明完毕。

这个定理可扩展运用到随机序列 $Y[k]$中，它是与下面随机序列 $X[k]$相关的鞅

对 $1<k\leqslant n$，有 $\mathcal{E}[Y[n]\mid X[0],\cdots,X[k-1]]=Y[k-1]$ (6-69)

为完整起见，我们给出如下定义。

定义(下鞅和上鞅) 随机序列 $X[k]$是**下鞅**，如果满足对所有 $k\geqslant 1$，有

$$\mathcal{E}[X[k]\mid X[0],\cdots,X[k-1]]\geqslant X[k-1] \tag{6-70}$$

如果满足对所有 $k\geqslant 1$，有

$$\mathcal{E}[X[k]\mid X[0],\cdots,X[k-1]]\leqslant X[k-1] \tag{6-71}$$

则是**上鞅**。

因为定义中包括了等号，所以鞅一定是上鞅和下鞅。注意到例 6-11 中伯努利序列在 $p>q$时是一个上鞅，而在 $p<q$ 时是一个下鞅(由于 $c>0$)。

最后，给出随机过程的鞅的定义，其中$\mathcal{T}$为连续时间。

定义(鞅过程) $\mathcal{E}[|X(t)|]<\infty$的随机过程 $X(t)$是一个鞅，如果满足对 $t_1<t_2$ 有

$$\mathcal{E}[X(t_2)\mid X(t),t\leqslant t_1]=X(t_1) \tag{6-72}$$

随机过程$Y(t)$，$\mathcal{E}[|Y(t)|]<\infty$，是一个相对于过程$X(t)$的鞅，如果满足对 $t_1<t_2$ 有

$$\mathcal{E}[Y(t_2)\mid X(t),t\leqslant t_1]=Y(t_1) \tag{6-73}$$

对一个连续时间鞅过程，期望值减为了在条件中只包含最近时间的随机变量。接下来的定理对独立增量过程非常重要，其中包括了几个非常著名的例子(有一些会在稍后讨论)

定理 6-5 一个均值为零的独立增量过程 $X(t)$是一个鞅

证明：根据均值运算的线性特性，可得：

$$\mathcal{E}[X(t_2)\mid X(t),t\leqslant t_1]=\mathcal{E}[X(t_2)-X(t_1)\mid X(t),t\leqslant t_1]+\mathcal{E}(X(t_1)\mid X(t),t\leqslant t_1] \tag{6-74}$$

由于此随机过程具有独立增量，等式右边的第一个均值运算的条件可以舍弃：

$$\mathcal{E}[X(t_2)\mid X(t),t\leqslant t_1]=\mathcal{E}[X(t_2)-X(t_1)]+\mathcal{E}[X(t_1)\mid X(t),t\leqslant t_1] \tag{6-75}$$

应用给定的零均值假设，得出：

$$\mathcal{E}[X(t_2)\mid X(t),t\leqslant t_1]=0+\mathcal{E}[X(t_1)\mid X(t),t\leqslant t_1]=X(t_1) \tag{6-76}$$

其中我们用到了等式右边的条件中包含 $X(t_1)$的这个事实，证明完毕。

注意到零均值的假设是必须的。

例 6-14 维纳过程 $W(t)$(稍后将讲到)是独立增量过程，它也是一个鞅。我们可以通过直接求条件均值来证明这个特性。稍后将在第 9 章中对联合高斯随机变量 X 和 Y 的参数估计中给出证明，条件均值为：

$$\mathcal{E}[X\mid Y]=\mu_X+(C_{XY}/\sigma_Y^2)(Y-\mu_Y) \tag{6-77}$$

对一个维纳过程，变量为 $X=X(t_2)$，$Y=X(t_1)$，$\mu_X=\mu_Y=0$，$C_{XY}(t_1,\ t_2)=\min(t_1,\ t_2)=t_1$ 以及 $\sigma_Y^2=t_1$。则马上可以得到：

对 $t_1<t_2$ 有 $\mathcal{E}[X(t_2)\mid X(t_1)]=X(t_1)$ (6-78)◄

6.9 马尔可夫序列

随机序列的以下特性包括了几个时刻的条件作用。

定义(马尔可夫序列) **马尔可夫序列** $X[k]$是一个时间索引的随机变量集，满足以下马尔可夫特性：

$$P(X[k]=x_k\mid X[1]=x_1,\cdots,X[k-1]=x_{k-1})=P(X[k]=x_k\mid X[k-1]=x_{k-1}) \tag{6-79}$$

马尔可夫序列通常也被称为马尔可夫链。

对一个马尔可夫序列，在之前所有时刻的输出都给定的条件下，某个特定时刻输出的概率只与最近时刻的输出有关。输出的完整历史是不需要的。式(6-79)中的特性有时被称为一步马尔可夫，它可以扩展到 m 步，如下所示。

定义(m 阶马尔可夫序列) 对 $k>m$，一个 m 阶马尔可夫序列具有如下性质：

$$P(X[k]=x_k|X[1]=x_1,\cdots,X[k-m-1]=x_{k-m-1},X[k-m]=x_{k-m},\cdots,X[k-1]=x_{k-1})$$
$$=P(X[k]=x_k|X[k-m]=x_{k-m},\cdots,X[k-1]=x_{k-1}) \quad (6\text{-}80)$$

尽管马尔可夫序列可以用可数无限个输出的集合来描述，但绝大多数马尔可夫应用包括的都是具有有限个输出的序列。这样的随机序列可用有限个状态来描述。

定义(马尔可夫链的状态) 马尔可夫链的状态是指在任意时刻 k 的随机变量 $X[k]$ 的可能输出。

为方便标记，我们将一步条件概率表示如下：

$$p_{mn}\triangleq P(X[k]=n|X[k-1]=m) \quad (6\text{-}81)$$

假设过程平稳，则$\{p_{mn}\}$不依赖特定的时刻值 k。状态通常都用$\mathbb{Z}^+$的子集做有序的编号，因此可以被看做由真实的输出到实数的映射，这样操作起来更方便。

假设有有限个状态，一个马尔可夫链可用一个状态转移矩阵来描述。设状态数为 M，将其标记为 $1,\cdots,M$，用来表述 $X[k]$的输出。列向量 $\boldsymbol{p}[k]\triangleq[p_1[k],\cdots,p_M[k]]^{\mathrm{T}}\in[0,1]^M$ 表示在时刻k 序列在M 种可能中的每一个状态的概率(注意不要将状态概率$p_m[k]$与概率质量函数(pmf)$p_X[x]$混淆)。状态转移矩阵 $\boldsymbol{P}\in[0,1]^{M\times M}$ 的元素为：

$$\boldsymbol{P}\triangleq\begin{bmatrix}p_{11} & \cdots & p_{1M}\\ \vdots & \cdots & \vdots\\ p_{M1} & \cdots & p_{MM}\end{bmatrix} \quad (6\text{-}82)$$

其中$\{p_{mn}\}$由式(6-81)给定。由条件概率的定义，考虑 $\boldsymbol{p}[k]$的第 n 个元素：

$$p_n[k]=\sum_{m=1}^{M}P(X[k]=n,X[k-1]=m)$$
$$=\sum_{m=1}^{M}P(X[k]=n|X[k-1]=m)P(X[k-1]=m) \quad (6\text{-}83)$$

求和运算作用于随机序列在 $k-1$ 时刻的所有输出(状态)。用 $\boldsymbol{P}$ 中元素的符号，我们得到：

$$p_n[k]=\sum_{m=1}^{M}p_{mn}p_m[k-1] \quad (6\text{-}84)$$

这是根据全概率公式得到的。对完整的状态向量，时刻 $k-1$ 转移到 k 满足以下递推关系：

$$\boldsymbol{p}^{\mathrm{T}}[k]=\boldsymbol{p}^{\mathrm{T}}[k-1]\boldsymbol{P} \quad (6\text{-}85)$$

也可写为列向量：$\boldsymbol{p}[k]=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{p}[k-1]$。以上行向量的表示方法主要用来与大多数有关马尔可夫链的书保持一致。这种表示方法更可取，是因为我们更习惯将 p_{mn} 看做是由状态 m 到 n(下标由左至右)，它需要将 $\boldsymbol{p}[k-1]$的元素用 $\boldsymbol{P}$ 的第 n 列的元素加权来得到 $p_n[k]$。

显然，$\boldsymbol{P}$ 的元素必须是非负的，因为它们是概率。如果有一个元素 p_{mn} 为零，则由状态 m 到 n 的转移就不会发生。但是只要 $p_{nm}\neq 0$，由状态 n 到 m 的反向转移就有可能。$\boldsymbol{P}$ 通常不是一个对称矩阵。从时刻 $k-1$ 到 k，必然要发生什么：或者移动到一个不同的状态，或者保持同一个状态。这意味着 $\boldsymbol{P}$ 的每一行的和为 1。总结如下，$\boldsymbol{P}$ 的两个最典型的性质为：

$$p_{mn}\geqslant 0,\qquad m,n=1,\cdots,M \quad (6\text{-}86)$$

$$\sum_{n=1}^{M}p_{mn}=1,\quad m=1,\cdots,M \quad (6\text{-}87)$$

状态转移矩阵通常也称为随机阵，由于每一行的和为 1，$\boldsymbol{P}$ 又被称为右随机阵。

例 6-15 图 6-8 给出了一个具有三个状态的马尔可夫链的状态转移图。设其具有如下状态转移矩阵：

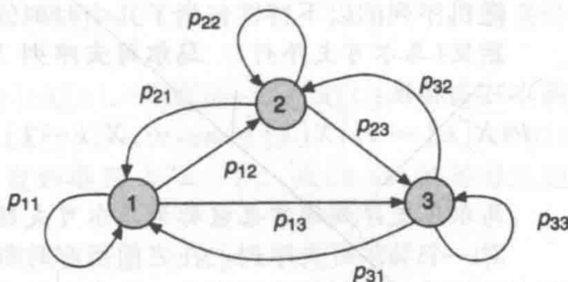


图 6-8 三个状态的马尔可夫链的例子

$$P=\begin{bmatrix}0.1 & 0.3 & 0.6\\0 & 0.5 & 0.5\\0.9 & 0.1 & 0\end{bmatrix} \tag{6-88}$$

当序列进入状态3，它一定会在下一时刻转移到一个不同的状态，因为 $p_{33}=0$.（对转移概率为零的情况，一般不会在状态转移图中标注箭头）。注意到矩阵中每一行的和都为1，列一般不满足（当然要看具体情况，一列或几列也可能和为1）。◀

我们总结一下马尔可夫链的一些重要性质：

定义（不可约） 一个马尔可夫链是**不可约**的，如果任何状态总可以由任意其他状态在某点时间到达。

这意味着存在 $k>0$ 满足：

$$P(X[k]=n|X[0]=m)>0 \tag{6-89}$$

不可约并不意味 $\boldsymbol{P}$ 中所有转移概率都是非零的。尽管例6-15中 $\boldsymbol{P}$ 中有两个元素为零，我们看到它仍可能到达任意状态（这点可以非常容易由状态转移图证明）。特别是，尽管序列不能由状态1转移到状态2，因为 $P_{12}=0$，但它仍可以通过先转移到状态3，再到达状态2。

例6-16 一个具有以下状态转移矩阵的马尔可夫链就不是不可约的：

$$P=\begin{bmatrix}0.5 & 0.5 & 0 & 0\\0.2 & 0.8 & 0 & 0\\0 & 0 & 0.1 & 0.9\\0 & 0 & 0.3 & 0.7\end{bmatrix} \tag{6-90}$$

状态1和2在任意时刻都不可能到达状态3和4，反之亦然。◀

这表明例6-16中的 $\boldsymbol{P}$ 有两个相通类。

定义（相通类） 马尔可夫链中的一个**相通类**是状态的一个子集，这些状态之间可以互相到达，与类之外的状态不能互通。

因此，一个不可约马尔可夫链只有一个相通类。对于例6-16，我们将两个相通类看做两个隔离的不可约马尔可夫链。

不同于一步转移，通常检查状态转移类型要方便些。考虑从状态 m 经过 r 步以如下条件概率转移到状态 n：

$$P(X[k]=n|X[k-r]=m)\triangleq p_{mn}^{(r)} \tag{6-91}$$

在定理6-6中，我们将证明怎样由低阶转移概率得到这些转移概率。

定理6-6(Chapman-Kolmogorov) 对马尔可夫链 $X[k]$，一个 r 步转移概率可由 $(r-1)$ 步转移概率计算得到，如下所示：

$$p_{mn}^{(r)}=\sum_{q=1}^{M}p_{mq}^{(r-1)}p_{qn} \tag{6-92}$$

其中 $p_{mq}^{(1)}\equiv p_{mq}$。

证明：由式(6-83)的全概率公式可得：

$$P(X[r]=n)=\sum_{q=1}^{M}P(X[r]=n|X[r-1]=q)P(X[r-1]=q) \tag{6-93}$$

以初始状态 $X[0]=m$ 为条件可得：

$$\begin{aligned}P(X[r]=n|X[0]=m)=&\sum_{q=1}^{M}P(X[r]=n|X[0]=m,X[r-1]=q)\\&\cdot P(X[r-1]=q|X[0]=m)\end{aligned} \tag{6-94}$$

用以前的符号，得到：

$$p_{mn}^{(r)}=\sum_{q=1}^{M}P(X[r]=n|X[0]=m,X[r-1]=q)p_{mq}^{(r-1)} \tag{6-95}$$

由于 $X[k]$ 是一个马尔可夫链，可以得到：

$$P(X[r]=n|X[0]=m,X[r-1]=q)=p(X[r]=n|X[r-1]=q)=p_{qn} \quad (6\text{-}96)$$

证明完毕。

定义 r 步状态转移矩阵 $\boldsymbol{P}^{(r)}$，其元素为 $\{p_{mn}^{(r)}\}$，由式(6-92)中的求和项给出如下递推关系：

$$\boldsymbol{P}^{(r)}=\boldsymbol{P}\boldsymbol{P}^{(r-1)} \quad (6\text{-}97)$$

反复应用 Chapman-Kolmogorov 等式，可得

$$\boldsymbol{P}^{(r)}=\boldsymbol{P}\boldsymbol{P}\boldsymbol{P}^{(r-2)}=\boldsymbol{P}^{r} \quad (6\text{-}98)$$

其中 $\boldsymbol{P}$ 是初始的一步转移矩阵。

定义(瞬态、常返态和吸收态)　**瞬态**是指如果进入此状态后，有一个非零的概率使得这个状态不会再次被到达。而**常返态**会以概率 1 再次到达。**吸收态**一旦到达，就不可能再离开这个状态。

显然对于吸收态 m，有 $p_{mm}=1$ 和 $p_{mn}=0$(当 $m\neq n$)。关于瞬态和常返态可用下式描述。定义条件概率：

$$\widetilde{p}_{n}^{(k)}\triangleq P(X[k]=n,X[m]\neq n,m=1,\cdots,k-1\,|\,X[0]=n) \quad (6\text{-}99)$$

这是序列只在 k 时刻回到初始状态 n 的概率，则：

$$\sum_{k=1}^{\infty}\widetilde{p}_{n}^{(k)}\triangleq\widetilde{p}_{n} \quad (6\text{-}100)$$

是序列在时间上某一点回到状态 n 的概率。因此，对常返态有 $\widetilde{p}_{n}=1$，而对瞬态有 $\widetilde{p}_{n}<1$。对可能或不可能再次到达的瞬态 $\widetilde{p}_{n}=0$ 不是必须的。这种不容易计算的表述导致了以下定理，我们将不加证明地给出定理的描述。

定理 6-7　状态 n 是瞬态，当且仅当

$$\sum_{r=1}^{\infty}p_{nn}^{(r)}<\infty \quad (6\text{-}101)$$

它是常返态，当且仅当

$$\sum_{r=1}^{\infty}p_{nn}^{(r)}=\infty \quad (6\text{-}102)$$

由此我们得到一个结论，我们只需检查转移矩阵 $\boldsymbol{P}$ 的对角线元素，就是一个马尔可夫链在下次转移时仍留在同一状态的概率。

例 6-17　考虑以下转移矩阵：

$$\boldsymbol{P}=\begin{bmatrix}0&0&0&1\\1/3&1/3&1/3&0\\0&0&1&0\\1&0&0&0\end{bmatrix} \quad (6\text{-}103)$$

这个简单的例子的特征很容易用图 6-9 中的状态转移图来描述。状态 3 很显然是一个吸收态，状态 1 和 4 是常返态。状态 2 是一个瞬态，因为它不会再次到达的概率为 2/3。◀

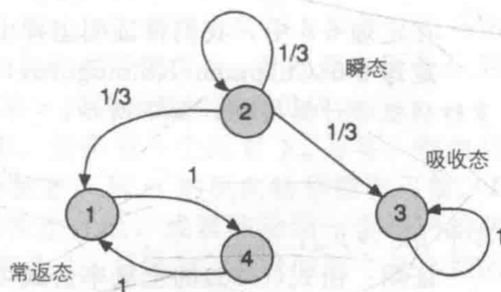


图 6-9　例 6-17 的状态转移图

利用式(6-98)的结果，只要初始状态 $\boldsymbol{p}[0]$ 已知，则未来任意时刻的状态概率都可能计算出。式(6-85)的递推形式对应于一个“输出”为 $\boldsymbol{p}[k]$ 的线性系统。它实际上是状态空间模型的一类，只是输入为零，并且写为行向量形式(见第 1 章)。我们知道这样一个线性系统的行为是受到 $\boldsymbol{P}$ 的特征值控制的，它与系统的极点一致。这些需要求解特征方程得到。

定义(特征方程)　一个转移矩阵为 $\boldsymbol{P}$ 的马尔可夫链的特征方程为：

$$\det(\boldsymbol{P}-\lambda\boldsymbol{I})=0 \quad (6\text{-}104)$$

其中 $\boldsymbol{I}\in\mathcal{R}^{M\times M}$ 是一个单位矩阵，λ 是一个标量，$\det(\boldsymbol{A})$ 是方阵 $\boldsymbol{A}$ 的行列式。

这个特征方程是一个 λ 的 M 阶多项式，最多有 M 个与 $\boldsymbol{P}$ 的特征值对应的独立解。

例 6-18 考虑以下二阶状态转移矩阵：

$$\boldsymbol{P}=\begin{bmatrix}0.2 & 0.8\\0.5 & 0.5\end{bmatrix} \tag{6-105}$$

特征方程为：

$$\det\begin{pmatrix}0.2-\lambda & 0.8\\0.5 & 0.5-\lambda\end{pmatrix}=\lambda^2-0.7\lambda-0.3 \tag{6-106}$$

$$=(\lambda-1)(\lambda+0.3)=0 \tag{6-107}$$

得到特征值 $[-0.3,1]$。◀

以下是 $\boldsymbol{P}$ 的一个有用的特性，因为它的每行的和为 1。

定理 6-8 当 $|\lambda|\leqslant 1$ 时，状态转移矩阵 $\boldsymbol{P}$ 的特征值是有界的。至少有一个特征值等于 1。

证明：由特征方程开始：

$$\boldsymbol{Pq}=\lambda\boldsymbol{q} \tag{6-108}$$

对非零右特征向量 $\boldsymbol{q}\in\mathcal{R}^M$。选择 $\boldsymbol{Pq}$ 的第 m 行，对 $1\leqslant n\leqslant M$，有 $|q_m|\geqslant|q_n|$：

$$\sum_{n=1}^{M}p_{mn}q_n=\lambda q_m \tag{6-109}$$

由于 $\boldsymbol{P}$ 中元素是非负的，我们得到以下的上界：

$$|\lambda q_m|\leqslant\sum_{n=1}^{M}p_{mn}|q_n| \tag{6-110}$$

由于 q_m 是 $\boldsymbol{q}$ 中所有元素中幅度最大的，求和运算中的每个 $|q_n|$ 可以用 $|q_m|$ 来替换以得到另一个上界，$|q_m|$ 可以分解为：

$$\sum_{n=1}^{M}p_{mn}|q_n|\leqslant|q_m|\sum_{n=1}^{M}p_{mn}=|q_m| \tag{6-111}$$

其中用到了 $\boldsymbol{P}$ 的每一行的和为 1(即对列索引 n 求和)。将这些结论联系起来得到：

$$|\lambda q_m|=|\lambda||q_m|\leqslant|q_m|\Rightarrow|\lambda|\leqslant 1 \tag{6-112}$$

注意到这个结论适用于 $\boldsymbol{P}$ 的任意特征值，上式中我们并没有选择一个特定的特征值 λ 和右特征向量 $\boldsymbol{q}$。最后，由于 $\boldsymbol{P}$ 的每一行的和为 1，可以得到：

$$\boldsymbol{P1}=\boldsymbol{1} \tag{6-113}$$

其中 $\boldsymbol{1}\triangleq[1,\cdots,1]^{\mathrm{T}}\in\mathcal{R}^M$。这是右向量为 $\boldsymbol{1}$，特征值 $\lambda=1$ 时的特征方程。因此，$\boldsymbol{P}$ 至少有一个特征值为 1，证明完毕。

假设 $\boldsymbol{P}$ 有 M 个线性独立的左特征向量，考虑 $\boldsymbol{P}$ 的以下分解形式：

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\Rightarrow\boldsymbol{P}=\boldsymbol{V}^{-\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{6-114}$$

其中 $\boldsymbol{V}^{\mathrm{T}}$ 的行是 $\boldsymbol{P}$ 的左特征向量，$\boldsymbol{\Lambda}$ 是一个包括了特征值 $\{\lambda_m\}$ 的对角线矩阵(见附录 G)。展开为乘积求和的形式得到：

$$\boldsymbol{P}=\sum_{m=1}^{M}\lambda_m\boldsymbol{s}_m\boldsymbol{v}_m^{\mathrm{T}} \tag{6-115}$$

其中 $\{\boldsymbol{v}_m^{\mathrm{T}}\}$ 是 $\boldsymbol{V}^{\mathrm{T}}$ 的行($\boldsymbol{V}$ 的列)，$\{\boldsymbol{s}_m\}$ 是 $\boldsymbol{V}^{-\mathrm{T}}$ 的列($\boldsymbol{V}^{-1}$ 的行)。注意到左、右特征向量是不相同的，除非 $\boldsymbol{P}$ 正好是对称的，通常马尔可夫链不满足这种情况。左特征向量并不是唯一的：$a\boldsymbol{v}_n^{\mathrm{T}}$(如果 $a\neq 0$)也是一个与 λ_n 相关的特征向量。为了方便起见，特征向量一般都是单位长度的：$\boldsymbol{v}_m^{\mathrm{T}}\boldsymbol{v}_m=1$。注意到，$\boldsymbol{P}^k$ 也可按上式分解，即 $\boldsymbol{P}^k$ 和 $\boldsymbol{P}$ 有相同的特征向量，唯一的差别是用 $\boldsymbol{\Lambda}^k$ 取代了 $\boldsymbol{\Lambda}$。通过做这样的特征分解，时刻 k 的状态概率可以被表示为如下所示的非递推形式：

$$\boldsymbol{p}^{\mathrm{T}}[k]=\sum_{m=1}^{M}\lambda_m^k\boldsymbol{p}^{\mathrm{T}}[0]\boldsymbol{s}_m\boldsymbol{v}_m^{\mathrm{T}}=\sum_{m=1}^{M}c_m\lambda_m^k\boldsymbol{v}_m^{\mathrm{T}} \tag{6-116}$$

其中 $c_m \triangleq \boldsymbol{p}^{\mathrm{T}}[0]\boldsymbol{s}_m$ 是一个标量。这个表达式具有等比级数的形式，因此为了确保 $\boldsymbol{p}[k]$ 随着 k 的增长是有限的，对任意 $m|\lambda_m|$ 必须不超过 1。尽管式(6-116)中需要 $\boldsymbol{P}$ 的左特征向量和特征值，我们也没有必要必须知道$\{\boldsymbol{s}_m\}$，而只需要知道$\{c_m\}$，在接下来的例题中会证明这些可由初始条件得到。最后，可以通过定义状态转移矩阵 $\boldsymbol{P}^{\mathrm{T}}$，将以上的结论重新用右向量的形式表示。但是，正如以前提到的，绝大部分书中关于马尔可夫链的 $\boldsymbol{P}$ 的定义都是 $P(X[k]=n|X[k-1]=m])=p_{mn}$。

状态的稳态概率，即 $\lim\limits_{k\to\infty}\boldsymbol{p}[k]=\boldsymbol{p}[\infty]$，可由式(6-116)中 $\lambda=1$ 的所有项得到，由于 $|\lambda|<1$ 时对应的其他项会衰减到零(例 6-20 中检查了 $\boldsymbol{P}$，它的两个特征值是单位幅值)。假设第 n 个特征值为 1，则稳态解为：

$$\boldsymbol{p}[\infty]=\lim_{k\to\infty}\sum_{m=1,m\neq n}^{M}\boldsymbol{c}_m\lambda_m^k\boldsymbol{v}_m+c_n\boldsymbol{v}_n=c_n\boldsymbol{v}_n \tag{6-117}$$

它与 $\lambda_n=1$ 时对应的特征向量成比例关系(尽管我们已将这个表达式写为了一个列向量，$\boldsymbol{v}_m^{\mathrm{T}}$ 仍是 $\boldsymbol{P}$ 的一个左特征向量。)

例 6-19 对例 6-18 中的 $\boldsymbol{P}$，特征向量可以通过解以下两个矩阵方程得到：

$$\boldsymbol{v}_1^{\mathrm{T}}(\boldsymbol{P}+0.3\boldsymbol{I})=\boldsymbol{v}_1^{\mathrm{T}}\begin{bmatrix}0.5 & 0.8\\0.5 & 0.8\end{bmatrix}=\boldsymbol{0} \tag{6-118}$$

$$\boldsymbol{v}_2^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{I})\boldsymbol{q}_2=\boldsymbol{v}_2^{\mathrm{T}}\begin{bmatrix}-0.8 & 0.8\\0.5 & -0.5\end{bmatrix}=\boldsymbol{0} \tag{6-119}$$

得到：

$$\boldsymbol{v}_1^{\mathrm{T}}=[1,\ -1],\ \boldsymbol{v}_2^{\mathrm{T}}=[1,\ 1.6] \tag{6-120}$$

$$\boldsymbol{p}[k]=c_1(-0.3)^k\begin{bmatrix}1\\-1\end{bmatrix}+c_2\begin{bmatrix}1\\1.6\end{bmatrix} \tag{6-121}$$

假设初始状态概率向量为 $\boldsymbol{p}[0]=[0.6,0.4]^{\mathrm{T}}$。由式(6-117)中稳态结果可知其和必为 1，则：

$$c_2+(1.6)c_2=1\Rightarrow c_2=1/2.6\approx 0.384 \tag{6-122}$$

由初始状态：

$$\boldsymbol{p}[0]=c_1\begin{bmatrix}1\\-1\end{bmatrix}+c_2\begin{bmatrix}1\\1.6\end{bmatrix}=\begin{bmatrix}0.6\\0.4\end{bmatrix} \tag{6-123}$$

第一行给出了 $c_1+c_2=0.6\Rightarrow c_1=0.6-1/2.6\approx 0.215$。我们也可以通过求解式(6-123)中两个线性方程得到 c_1 和 c_2 的值。当然，就如上题所做的，通常情况下首先由分解式的稳态项求出一个系数的方法简单些(因为至少有一个特征值为 1)。在 $M=3$ 时，这种方法更有用，因为一旦一个系数已知，很容易通过解三个线性方程中的两个来得到另两个系数。◀

通常，$\boldsymbol{P}$ 是一个非负矩阵。

定义(正阵和非负矩阵) 如果矩阵 $\boldsymbol{A}$ 中所有元素严格为正，则矩阵 A 为**正阵**。如果所有元素非负，则它为**非负矩阵**。

正阵与正定矩阵是不同的，正定矩阵中所有的特征值严格为正。而我们在例 6-18 中看到正阵 $\boldsymbol{P}$，它的特征值可以为负。如果 $\boldsymbol{P}$ 正好是正阵，则以下的定理描述了一些非常有用的特性，我们不加证明地给出。

定理 6-9(佩龙-弗罗宾尼斯(Perron-Frobenius)) 对于正阵 $\boldsymbol{A}\in(0,\infty)^{M\times M}$

- 存在一个实数特征值 $\lambda_0>0$，这样所有其他特征值满足 $|\lambda|<\lambda_0$。
- λ_0 的重数为 1。
- 存在一个与 λ_0 相关的左特征向量 $\boldsymbol{v}$ 和右特征向量 $\boldsymbol{q}$：$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}=\lambda_0\boldsymbol{v}^{\mathrm{T}}$ 和 $\boldsymbol{A}\boldsymbol{q}=\lambda_0\boldsymbol{q}$，$\boldsymbol{v}$ 和 $\boldsymbol{q}$ 只有正的元素。
- 不存在其他正的特征向量。
- λ_0 的大小受 $\boldsymbol{A}$ 中元素约束，具体关系如下：

$$\min_m \sum_{n=1}^{M} a_{mn} \leqslant \lambda_0 \leqslant \max_m \sum_{n=1}^{M} a_{mn} \tag{6-124}$$

最后一个特性确保了对于一个马尔可夫链的状态转移矩阵 $\boldsymbol{P}$ 有 $\lambda_0=1$。如果 $\boldsymbol{P}$ 只是非负的，则不能采用这个定理的所有结论。用 $|\lambda|\leqslant\lambda_0$ 来代替（等号也可以成立），相关的左和右特征向量是非负的。

例 6-20 考虑以下非负转移矩阵：

$$\boldsymbol{P}=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \tag{6-125}$$

由 $\det(\boldsymbol{P}-\lambda\boldsymbol{I})=\lambda^2-1=0$ 可得特征值为 ±1。对这个 $\boldsymbol{P}$ 为非负的简单的例子，Perron-Frobenius 定理的第一个结论并不满足严格不等式。与 $\lambda=\{-1,1\}$ 对应的特征向量分别为 $(1/\sqrt{2})[-1,1]^{\mathrm{T}}$ 和 $(1/\sqrt{2})[1,1]^{\mathrm{T}}$。对这个例子，式(6-116)的表达式为：

$$\boldsymbol{p}[k]=\frac{1}{\sqrt{2}}\left[c_1(-1)^k\begin{bmatrix}-1\\ 1\end{bmatrix}+c_2\begin{bmatrix}1\\ 1\end{bmatrix}\right] \tag{6-126}$$

因为 $\lambda=-1$，则其不存在一个稳态。这个序列不停地在状态 1 和 2 之间转换。◀

例 6-21 考虑正阵：

$$\boldsymbol{P}=\begin{bmatrix}1/3 & 1/3 & 1/3\\ 1/4 & 1/2 & 1/4\\ 2/3 & 1/6 & 1/6\end{bmatrix} \tag{6-127}$$

式(6-116)为：

$$\boldsymbol{p}[k]=c_1\begin{bmatrix}0.669\\ 0.595\\ 0.446\end{bmatrix}+c_2(-0.204)^k\begin{bmatrix}0.798\\ -0.247\\ -0.550\end{bmatrix}+c_3(0.204)^k\begin{bmatrix}-0.603\\ 0.778\\ -0.175\end{bmatrix} \tag{6-128}$$

与 $\lambda=1$ 对应的特征向量是唯一的正特征向量；对此例中 $\boldsymbol{P}$，佩龙-弗罗宾尼斯定理中所有结论都成立。稳态结果为：

$$\boldsymbol{p}[\infty]=c_1\begin{bmatrix}0.669\\ 0.595\\ 0.446\end{bmatrix} \tag{6-129}$$

$c_1=0.585$，则向量中元素的和为 1。◀

例 6-22 在这个例子中，我们考虑一个大家熟知的马尔可夫链：一维随机游走（如图 6-10所示）。存在 $N+1$ 种状态，转移只在相邻状态间发生，或者保持在当前状态。转移矩阵 $\boldsymbol{P}\in\mathcal{R}^{(N+1)\times(N+1)}$ 具有如下形式：

$$\boldsymbol{P}=\begin{bmatrix}r_0 & p_0 & 0 & \cdots & & 0\\ q & r & p & & & \vdots\\ 0 & q & \ddots & \ddots & & \vdots\\ \vdots & & \ddots & & p & 0\\ \vdots & & & & r & p\\ 0 & \cdots & \cdots & 0 & q_N & r_N\end{bmatrix} \tag{6-130}$$

其为三对角线的。注意到对内部状态满足 $p+q+r=1$，但对状态 0 和 N 分别满足 $r_0+p_0=1$ 和 $q_N+r_N=1$。这样一个转移矩阵会出现在有两个玩家 A 和 B 的赌博竞赛中，其中最大的赢数为 N。设图 6-10 代表的状态 $X[k]$ 是 A 在时刻 k 赢的数目，显然，B 赢的数目为 $N-X[k]$。这种竞赛可以用投掷一枚有偏的硬币来模拟，设 $P(H)=p$ 为 A 赢的概率，则 B 赢的概率为 $q\triangleq1-p$。对投掷硬

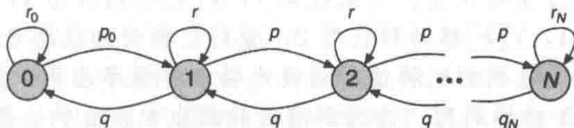


图 6-10 例 6-22 中一维随机游走的状态转移图

币竞赛，$r=0$(因为每次投掷一定会有人赢)，当一个玩家累积赢的数目达到 N 时，竞赛结束。这意味着状态 0 和 N 是吸收态：$r_0=r_N=1$。假设初始条件为 $X[0]=n$：玩家 A 处于状态 n，这样玩家 B 赢的数目为 $N-n$。玩家 A 输的概率为 $P(X[K]=0\mid X[0]=n)$，其中 $K\triangleq\min\{k\geqslant 0:\ X[k]=0$ 或 $X[k]=N)\}$ 为竞赛持续期的一个随机变量。可以表示如下(见例 6-26)：

$$P(X[K]=0\mid X[0]=n)=\begin{cases}(p^{N-n}q^{n}-q^{N})/(p^{N}-q^{N}), & p\neq q\\ (N-n)/N, & p=q=1/2\end{cases}\tag{6-131}$$

很有意思的是当 p 仅比 1/2 稍微少一点时，A 输的概率却会急剧增加，这是因为式(6-131)中的非线性形式。图 6-11 是 $N=20$ 的情况。这些图不是 pmf：纵轴对应的是式(6-131)中特定的条件概率。

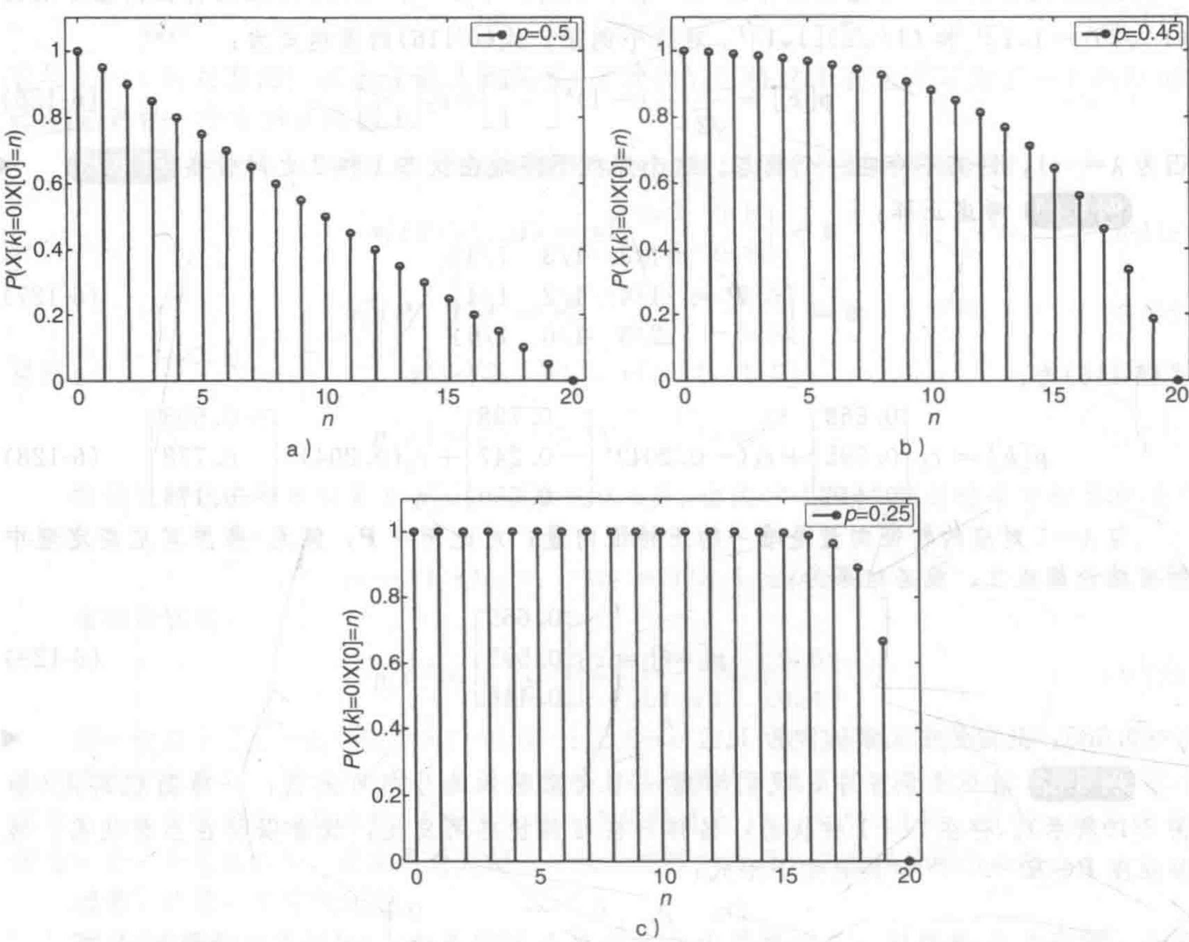


图 6-11　玩家 A 在例 6-22 中输的概率，其中初始赢 n 次，总共赢 $N=20$ 次。a) $p=0.5$；b) $p=0.45$；c) $p=0.25$ ◀

例 6-23　稍后描述的伯努利序列是数字通信系统中比特序列建模的有用模型。假设我们对计算特定比特样式感兴趣，比如输出为 $\{0,1\}$ 的伯努利序列 $X[k]$第 1 次出现 1，1。为了检测这种样式，我们可以构造一个有限状态机(FSM)(用于数字逻辑设计中的术语)，它等价于一个离散时间的马尔可夫链。设 $Y[k]$为 k 时刻的状态。这个链中需要三个状态。设状态 0 为初始状态。如果收到 1，$Y[k]$移到状态 1；否则它仍留在状态 0。如果处在状态 1 时，收到 1，$Y[k]$移动到状态 2；否则它移动到状态 0。状态转移图显示在图 6-12 中，其中转移标注的是收到的比特值。假设比特值等概率出现，转移概率都是 1/2。当到达状态 3，比特序列 1，1 被检测到。注意到虚线转移也包括在内，用来显示感兴趣的所有非叠加 1，1 比特序列(而不仅仅是第一次发生)的状态图。在以实线标注的状态转移图中，转移矩阵为：

$$\boldsymbol{P}=\begin{bmatrix}1/2 & 1/2 & 0\\ 1/2 & 0 & 1/2\\ 0 & 0 & 1\end{bmatrix} \tag{6-132}$$

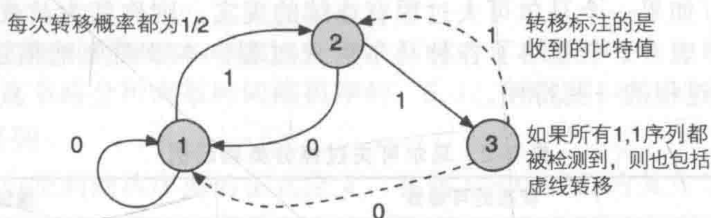


图 6-12　状态转换图

其中包括 $p_{33}=1$，只用来表明状态 3 是一个吸收态。当然，一旦到达状态 3，本例中将不再检测未来的比特值。◀

假设我们想知道到达状态 3 需要的平均比特数 N。可以通过大家熟知的一步分析法来确定，它是以所有的状态转移为条件。考虑以下条件期望：

$$n_1 \triangleq \mathcal{E}[N|Y[0]=1], n_2 \triangleq \mathcal{E}[N|Y[0]=2], n_3 \triangleq \mathcal{E}[N|Y[0]=3] \tag{6-133}$$

其中 n_m 是给定 $Y[0]$处在状态 m，到达状态 3 所需的平均比特数。显然，$n_3=0$；另外两种情况可由状态转移图的马尔可夫性得到。首先观察到：

$$\mathcal{E}[N|Y[0]=1]=1+\mathcal{E}[N-1|Y[0]=1] \tag{6-134}$$

$$\mathcal{E}[N|Y[0]=2]=1+\mathcal{E}[N-1|Y[0]=2] \tag{6-135}$$

其中每个表达式中第一个 1 只是 N 和 $N-1$ 之间的计数差。由全概率定理可得：

$$n_1=1+\sum_{m=1}^{3}\mathcal{E}[N-1|Y[1]=m]P(Y[1]=m|Y[0]=1) \tag{6-136}$$

$$n_2=1+\sum_{m=1}^{3}\mathcal{E}[N-1|Y[1]=m]P(Y[1]=m|Y[0]=2) \tag{6-137}$$

我们以"第一步"为条件(例如，状态转移从 $Y[0]$到 $Y[1]$)。尽管我们想找到 n_1，也就是以状态 1 开始到达状态 3 需要的比特数，一步分析需要上面的方程也同时解出。关键在于识别出平稳性的原因：

$$\mathcal{E}[N-1|Y[1]=m]=\mathcal{E}[N|Y[0]-m] \tag{6-138}$$

由 $Y[0]$移动到 $Y[1]$需要 1 比特。将状态转移矩阵($n_3=0$)中的概率代入得到：

$$n_1=1+(1/2)n_1+(1/2)n_2, n_2=1+(1/2)n_1 \tag{6-139}$$

因此平均的比特数为 $n_1=6$。从状态 2 起始到达状态 3 需要的平均比特数也可以得到：$n_2=4$。

6.10　马尔可夫过程

考虑连续时间随机过程 $X(t)$在几个不同时刻 $t_1<\cdots<t_{N-1}<t_N$ 对应的随机变量，我们可以得到联合 pdf $f_{X(t_1),\cdots,X(t_N)}(x(t_1),\cdots,x(t_N))$。通常，过程的不同时刻对应的随机变量是相关的，或者可能是独立的，这种情况下联合 pdf 分解为 N 个边缘 pdf$\{f_{X(t_k)}(x(t_k))\}$ 的乘积。我们感兴趣的是连续时间随机过程的马尔可夫性。

定义(马尔可夫过程)　随机过程 $X(t)$是**马尔可夫过程**，如果条件 pdf 可以简化为：

$$f_{X(t_N)|X(t_1),\cdots,X(t_{N-1})}(x(t_N)|x(t_1),\cdots,x(t_{N-1}))=f_{X(t_N)|X(t_{N-1})}(x(t_N)|x(t_{N-1})) \tag{6-140}$$

马尔可夫性也可以用条件 cdf 表示为：

$$\begin{aligned}P(X(t_N)\leqslant x(t_N)|X(t_1)&=x(t_1),\cdots,X(t_{N-1})=x(t_{N-1}))\\&=P(X(t_N)\leqslant x(t_N)|X(t_{N-1})=x(t_{N-1}))\end{aligned} \tag{6-141}$$

对于离散情况，条件可以简化到只需包括条件中定义的最近时刻的随机变量。在时间 t_{N-1}之前的随机变量都不相关。

马尔可夫过程一词常用来描述具有以上定义中特性的任意过程，适用于连续或离散时间。如果过程具有有限或可数个状态，则通常被称为马尔可夫链。6.9 节中马尔可夫链的内容只包括对离散时间的分析，但连续时间的马尔可夫链也是可能的(稍后有一个例子是关于泊松过程的)。如果一个马尔可夫过程有连续的现实，则称其为扩散过程(稍后有一个例子是维纳过程)。表 6-2 中总结了各种马尔可夫过程。本节简要地描述具有连续现实的连续时间马尔可夫过程的一些特性。

表 6-2 马尔可夫过程分类和示例

	状态的可数数	连续现实
离散时间 k	马尔可夫链(有限状态机)	无
连续时间 t	连续时间马尔可夫链(泊松过程)	扩散过程(维纳过程)

切普曼-柯尔莫哥洛夫(Chapman-Kolmogorov)方程也适用于连续时间过程。

定理 6-10(切普曼-柯尔莫哥洛夫方程) 对马尔可夫随机过程 $X(t)$

$$f_{X(t_3)|X(t_1)}(x(t_3)|x(t_1))=\int_{-\infty}^{\infty}f_{X(t_3)|X(t_2)}(x(t_3)|x(t_2))f_{X(t_2)|X(t_1)}(x(t_2)|x(t_1))\mathrm{d}x(t_2) \tag{6-142}$$

其中 $t_1<t_2<t_3$。

证明：回忆一下边缘 pdf 可通过对联合 pdf 积分得到，即：

$$\begin{aligned}f_{X(t_3)}(x(t_3))&=\int_{-\infty}^{\infty}f_{X(t_2),X(t_3)}(x(t_2),x(t_3))\mathrm{d}x(t_2)\\&=\int_{-\infty}^{\infty}f_{X(t_3)|X(t_2)}(x(t_3)|x(t_2))f_{X(t_2)}(x(t_2))\mathrm{d}x(t_2)\end{aligned} \tag{6-143}$$

其中第二行用条件概率替换。如果以 $X(t_1)$为条件得到：

$$f_{X(t_3)|X(t_1)}(x(t_3)|x(t_1))=\int_{-\infty}^{\infty}f_{X(t_3)|X(t_1),X(t_2)}(x(t_3)|x(t_1),x(t_2))f_{X(t_2)|X(t_1)}(x(t_2)|x(t_1))\mathrm{d}x(t_2) \tag{6-144}$$

由于 $X(t)$是一个马尔可夫过程，我们得到：

$$f_{X(t_3)|X(t_1),X(t_2)}(x(t_3)|x(t_1),x(t_2))=f_{X(t_3)|X(t_2)}(x(t_3)|x(t_2)) \tag{6-145}$$

证明完毕。

定理 6-11 如果 $X(t)$是一个独立增量随机过程，则其也是一个马尔可夫过程。

证明：考虑 $X(t)$在三个时刻 $t_1<t_2<t_3$ 的条件 pdf $f_{X(t_3)|X(t_1),X(t_2)}(x(t_3)|x(t_1),x(t_2))$，对于增量 $X(t_3)-X(t_2)$，我们可以得到：

$$f_{X(t_3)|X(t_1),X(t_2)}(x(t_3)|x(t_1),x(t_2))=f_{X(t_3)-X(t_2)|X(t_1),X(t_2)}(x(t_3)-x(t_2)|x(t_1),x(t_2)) \tag{6-146}$$

等式右边可用第 4 章中线性变换 $X(t_3)\rightarrow X(t_3)-X(t_2)$的变换方法得到。很容易得到上面的关系，因为以 $X(t_2)$为条件则意味着从 $X(t_3)$中减去一个常数。同样地，由于条件中还包括 $X(t_1)$，另一个线性变换 $X(t_2)\rightarrow X(t_2)-X(t_1)$，则得到：

$$\begin{aligned}&f_{X(t_3)-X(t_2)|X(t_1),X(t_2)}(x(t_3)-x(t_2)|x(t_1),x(t_2))\\&=f_{X(t_3)-X(t_2)|X(t_2)-X(t_1),X(t_2)}(x(t_3)-x(t_2)|x(t_2)-x(t_1),x(t_2))\end{aligned} \tag{6-147}$$

由于条件的作用，初始条件的 pdf 形式没有改变，只是发生了转换。因为 $X(t)$是一个独立增量过程，$X(t_3)-X(t_2)$和 $X(t_2)-X(t_1)$是独立的，因此 $X(t_2)-X(t_1)$可从条件中去掉：

$$\begin{aligned}&f_{X(t_3)-X(t_2)|X(t_2)-X(t_1),X(t_2)}(x(t_3)-x(t_2)|x(t_2)-x(t_1),x(t_2))\\&=f_{X(t_3)-X(t_2)|X(t_2)}(x(t_3)-x(t_2)|x(t_2))\end{aligned} \tag{6-148}$$

最后，由于 $X(t_2)$是条件变量，$X(t_3)-X(t_2)$可用 $X(t_3)$代替，则得到另一个变换，因此：

$$f_{X(t_3)-X(t_2)|X(t_2)}(x(t_3)-x(t_2)\,|\,x(t_2)) = f_{X(t_3)|X(t_2)}(x(t_3)\,|\,x(t_2)) \tag{6-149}$$

证明完毕。

6.11 随机序列

我们已经描述了随机过程的一些基本性质，并且对具有特殊分布和特性的重要的随机过程进行了分析。这节将分析离散时间随机序列，6.12 节将关注连续时间随机过程。

6.11.1 伯努利序列

我们首先给出伯努利随机序列的正式定义，在例 6-6 中我们将其作为重复投掷硬币的模型

定义（伯努利序列） **伯努利序列** $X[k]$是输出为 $\{0,1\}$ 的 iid 随机变量$\{X[0], X[1], \cdots\}$的集合，其中 $P(X[k]=1)=p$，$P[X[k]=0)=q\triangleq 1-p$。

这个序列严格平稳。每个随机变量都可以被看做是具有两个输出的一次试验。输出 1 指一次成功，而 0 则意味着一次失败。伯努利序列的期望为$\mathcal{E}[X[k]]=p$，二阶矩为$\mathcal{E}[X^2[k]]=p$。方差为 $\mathrm{var}[X[k]]=p-p^2=pq$。当然，对伯努利随机变量也会得到同样的结果，因为根据上面的定义 pdf 不会随时间的改变而改变。

我们还可以回顾一下第 3 章中讨论的用随机变量来建模的一些相关事件。设 N_1 为在 N 次试验中成功的次数。则 N_1 为二项随机变量，其 pmf 为：

$$P(N_1=n)=\binom{N}{n}p^n q^{N-n} I_{\{0,\cdots,N\}}[n] \tag{6-150}$$

设 N_2 为成功 $r(r\in\mathcal{Z}^+)$ 次所需的试验次数。则 N_2 为负二项随机变量，其 pmf 为：

$$P(N_2=n)=\binom{n+r-1}{r-1}p^r q^n I_{\mathcal{Z}^+}[n] \tag{6-151}$$

当 $r=1$ 时得到 N_2 的特例，从而得到定义域为$\mathcal{Z}^+$的几何随机变量。

如果输出变为 $\{-1,1\}$，则 $X[k]$是一个对称的伯努利序列。当 $p=q=1/2$ 时，图 6-13 显示了这个序列的一个现实。

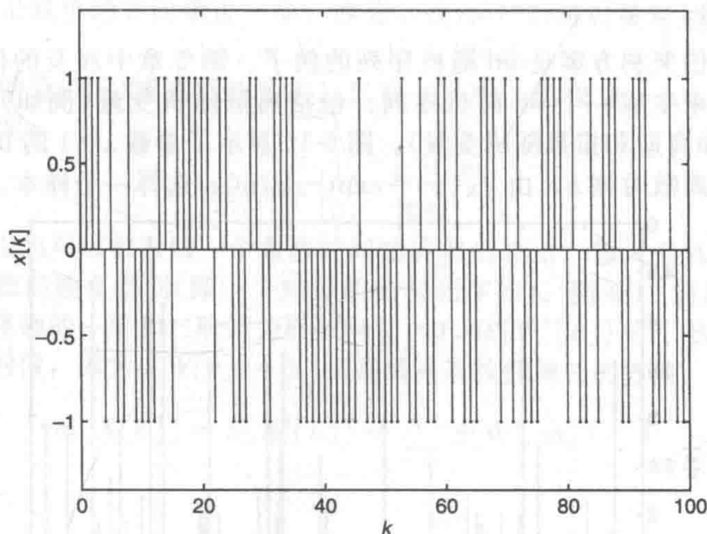


图 6-13　$p=0.5$ 时对称伯努利随机序列的现实 $x[k]$（竖线会使看起来容易些）

6.11.2 伯努利方案

伯努利方案即在每个时刻都有不止两个输出的伯努利序列的一种广义版本。

定义（伯努利方案） **伯努利方案** $X[k]$是具有离散输出 $\{1,\cdots,N\}$ 的 iid 随机变量 $\{X[0],X[1],\cdots\}$ 的集合，则 $P(X[k]=n)\triangleq p_n$，$\sum_{n=1}^{N} p_n=1$。

伯努利方案是严格平稳的，因为随机变量是 iid 的。连续以概率 $P(X[k]=n)=1/6$ $(n=1,\cdots,6)$ 掷筛子是一个伯努利方案的例子。对 $N>2$，随机变量的 pdf 理论上并不是伯努利，而且它也不一定要如一般掷筛子试验中满足均匀分布。它被称为伯努利方案仅仅因为它是一个伯努利序列到多个输出的自然扩展。图 6-14 中给出了一个具有四个等概率输出的伯努利方案的一个现实。在数字通信中使用多电平伯努利方案做发射模型的情况将在第 10 章中讨论，这被称为脉冲幅度调制(PAM)。对这样的系统，输出可以扩展到由脉冲形状所确定的连续时间内某个固定时期，这样就得到了一个随机过程 $X(t)$。稍后图 6-29显示的就是一个矩形脉冲情况下的一个现实。

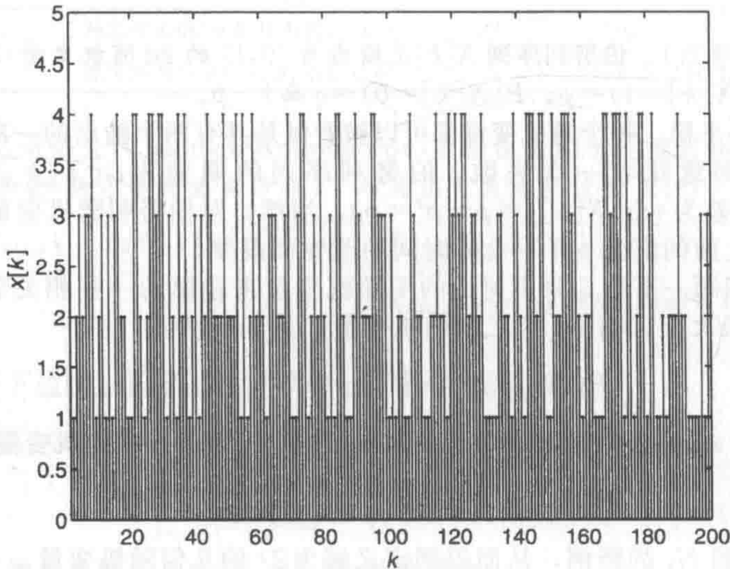


图 6-14 均匀分布的伯努利方案的一个现实 $x[k]$，输出为 $\{1,2,3,4\}$（竖线会使看起来容易些）

6.11.3 独立序列

伯努利序列和伯努利方案是 iid 随机序列的例子。第 3 章中涉及的任意随机变量基本上都可由时间索引来生成一个 iid 随机序列。包括离散随机变量(例如几何和泊松变量)、连续随机变量(例如高斯和指数随机变量)。图 6-15 显示了参数 $\lambda=1$ 的 iid 指数随机序列的一个现实，在每个离散时刻 k，由 $f_X(x)=\exp(-x)u(x)$ 选择一个样本。

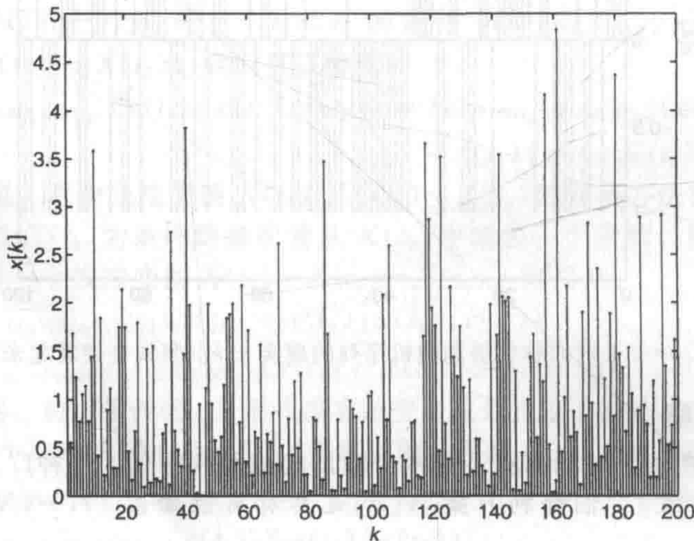


图 6-15 $\lambda=1$ 的独立同分布指数序列的一个现实 $x[k]$(竖线会使看起来容易些)

如果取消同分布的条件，则$\{X[k]\}$独立，但不必对每个k具有相同的分布。一个简单的例子是允许之前的指数序列的参数变化。例如，对于$k\in\mathcal{Z}^+$，假定$\lambda[k]=2+4/(k+1)$。每个随机变量仍为指数，但参数从最大的6变化到最小的2。在第7章，我们定义了当$k\to\infty$时序列的不同收敛类型。这个简单的例子是一个随机序列，该序列依分布收敛于参数$\lambda=2$的指数随机变量。

一个独立的随机序列，也许是序列的最基本类型，可以由它产生许多其他类型的随机序列。例如，伯努利随机游走和下面描述的二项式计数序列，它是对伯努利随机序列的输出求和生成的。这种求和对应于一类滤波器，作用于伯努利的输入，产生一个不再独立的输出序列。如图6-16所示的这种类型的模型，被广泛用于描述随机序列，将在第8章中讨论。该滤波器可以是线性的或非线性的，也可以是时变或时不变的。图6-1和引言中描述的现实是由一个线性时不变滤波器处理独立的高斯序列生成的。在第7章中，我们描述了为线性系统建模的常系数微分方程。这样的系统可以通过一个冲激响应函数来表示，在频域则可以用传递函数描述。

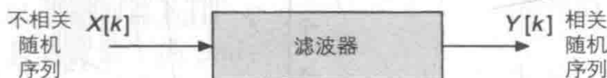


图6-16　滤波器作用于一个iid序列

6.11.4　伯努利随机游走

我们再回到伯努利序列，描述一个既不独立也不满足同分布的“随机游走”。

定义(伯努利随机游走)　设$W[k]$是一个输出为$\{-1,1\}$的对称伯努利序列，满足$P(W[k]=1)=p=P(W[k]=-1)=q=1/2$。随机序列：

$$X[k]=\sum_{m=1}^{k}W[m] \tag{6-152}$$

其中$X[0]=W[0]=0$是一个**伯努利随机游走**。

显然，和序列$X[k]$不是伯努利：随着k的增长，它可以是任何整数Z。伯努利随机游走的名字来源于式(6-152)中构成和的基本随机变量的pdf。它之所以被称为随机游走是因为序列随机地在正或负的方向前进一步。注意，式(6-152)可以被写做：

$$X[k]=X[k-1]+W[k] \tag{6-153}$$

它具有无限长冲激响应(IIR)滤波器的形式，只有一个唯一的极点$z=1$。冲激响应可由离散单位阶跃函数给定：

$$u[k]\triangleq\begin{cases}1, & k\in\mathcal{Z}^+\\ 0, & \text{其他}\end{cases} \tag{6-154}$$

伯努利随机游走也可以被看做一个离散时间滤波器的输出，输入是具有随机振幅$\{-1,1\}$的克罗内克δ函数的独立序列(即一个对称的伯努利序列)。图6-17给出了一个现实的例子。基本的伯努利序列的一阶和二阶矩为$\mathcal{E}[W[k]]=0$和$\mathcal{E}[W^2[k]]=1$。因此$X[k]$随时间逐步形成以0为中心的值，因为$\mathcal{E}[W[k]]=0$。但是其具有线性增长的方差：

$$\text{var}[X[k]]=\mathcal{E}[X^2[k]]=\sum_{m=1}^{k}\mathcal{E}[W^2[m]]=k \tag{6-155}$$

我们还观察到：

$$X[k+N]-X[k]=\sum_{m=k+1}^{k+N}W[m] \tag{6-156}$$

与$X[k]$独立。对于$X[k]$任意一组非重叠的区间，独立性都会保持。因此，伯努利随机游走是isi序列的一个例子。但是$X[k]$本身却不平稳，因为它的方差随时间而改变。

尽管伯努利随机游走在理论上定义$p=q=1/2$，但它可以推广为$p\neq q$。对这种情况，游走不再以0为中心，而且均值也是时变的：

$$\mathcal{E}[X[k]]=\sum_{m=1}^{k}\mathcal{E}[W[m]]=(p-q)k \tag{6-157}$$

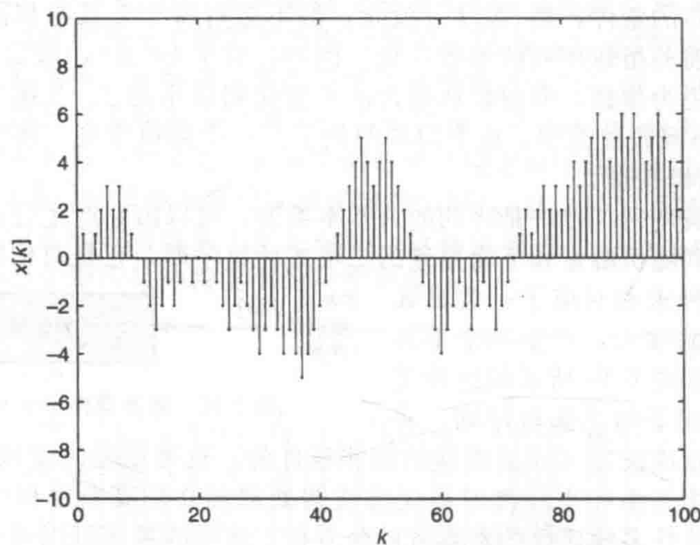


图 6-17　$p=0.5$ 的一个伯努利随机游走的现实示例

它或者增长($p>q$)或者衰减($p<q$)，因为当输出为 $\{-1,1\}$ 时$\mathcal{E}[W[m]]=p-q$，二阶矩为：

$$\mathcal{E}[X^2[k]] = k\,\mathcal{E}[W^2[k]] + (k^2-k)(\mathcal{E}[W[k]])^2 = k + (p-q)^2(k^2-k) \tag{6-158}$$

由于 $1-(p-q)^2=4pq$，方差为一个 k 的递增函数：

$$\text{var}[X(k)] = 4pqk \tag{6-159}$$

自相关函数和自协方差函数在例 6-13 中推导。

6.11.5　二项式计数序列

再次考虑输出为 $\{0,1\}$ 的伯努利随机序列。

定义(二项式计数序列)　设 $W[k]$ 为一个伯努利随机序列，$P(W[k]=1)=p$，$P(W[k]=0)=q$。随机序列

$$X[k] = \sum_{m=1}^{k} W[m] \tag{6-160}$$

其中 $X[0]=W[0]=0$，为一个**二项式计数序列**。

与之前的随机游走不同，二项式计数序列不会衰减。这当然是因为输出中 0 代替了 −1。由(6-160)式中的求和以及定理 6-3，我们发现二项式计数序列是另一个 isi 序列的例子。对一个确定的 k 值，随机变量 $X[k]$具有二项式 pmf：

$$P(X[k]=n) = \binom{k}{n} p^n q^{k-n} I_{\{0,\cdots,k\}}[n] \tag{6-161}$$

二项式计数序列也可以由式 6-153 中的 IIR 滤波器的输出来表示，滤波器冲激响应由离散单位阶跃函数给定。这里的主要不同在于，输入序列的克罗内克 δ 函数不均匀的时间间隔，因为伯努利的输出是 $\{0,1\}$。因此，某个现实的振幅会因为一串伯努利零的出现而在几个时间间隔保持常数。稍后讨论的泊松计数过程，也会出现相似的情况，这与连续时间二项式计数过程对应。表 6-3 中给出了基于伯努利序列的两个随机游走和计数序列的总结。

表 6-3　基于伯努利序列 $W[k]$的随机序列

随机序列 $X[k]=\sum_{m=1}^{k} W[m]$	$W[m]$的输出	参数
伯努利随机游走	$\{-1,1\}$	$p=1/2$
广义伯努利随机游走	$\{-1,1\}$	$0<p<1$
二项式计数序列	$\{0,1\}$	$0<p<1$

接下来，我们研究二项式计数序列中连续的伯努利 1 到达的间隔时间。由于计数过程是isi，可对$\mathcal{T}$中的任意间隔进行检查。为方便起见，检查第一个 1 的出现，我们用随机变量 Y 表示。做这样的选择是因为事件$\{Y>t\}$等价于事件$\{X(t)=0\}$。对于 1 的第一次出现是时刻 k 之后的情况，我们可以写作：

$$P(Y>k)=P(X[k]=0)=\binom{k}{0}q^k=q^k \tag{6-162}$$

其中用到了式(6-161)。则有：

$$P(Y\leqslant k)=(1-q^k)I_{\mathcal{N}}[k] \tag{6-163}$$

这对应于定义域为$\mathcal{N}$的几何随机变量的 cdf，且具有以下的 pmf：

$$p_Y[k]=p(1-q)^{k-1}I_{\mathcal{N}}[k] \tag{6-164}$$

（注意到这是附录 A 中的位移几何分布）。

二项式计数序列的均值为$\mathcal{E}[X[k]]=pk$，二阶矩为：

$$\begin{aligned}\mathcal{E}[X^2[k]]&=k\,\mathcal{E}[W^2[k]]+(k^2-k)(\mathcal{E}[W[k]])^2\\&=pk+p^2(k^2-k)\end{aligned} \tag{6-165}$$

方差为 $\mathrm{var}(X[k])=pk-p^2k=pqk$。图 6-18 给出了一个现实的例子。它的自相关函数为：

$$R_{XX}[k_1,k_2]=\mathcal{E}\Big[\sum_{m=1}^{k_1}W[m]\sum_{n=1}^{k_2}W[n]\Big] \tag{6-166}$$

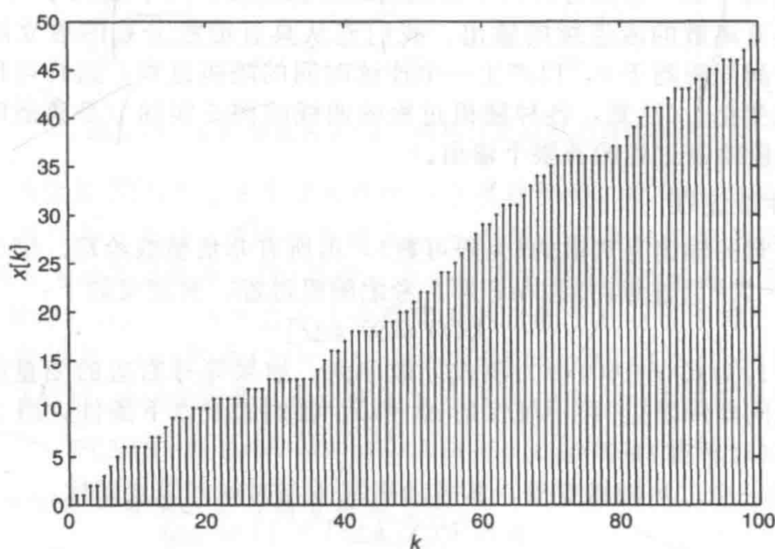


图 6-18 $p=0.5$，输出为 $\{0,1\}$ 的二项式计数序列的现实示例

为了继续往下讨论，假设 $k_2>k_1$；相反的情况可由对称性导出。由于$\{W[k]\}$是独立的，每个交叉项的期望值是各期望值的乘积：

$$\begin{aligned}R_{XX}[k_1,k_2]&=\sum_{m=1}^{k_1}\Big(\mathcal{E}[W^2[m]]+\sum_{n=1,n\neq m}^{k_2}\mathcal{E}[W[m]]\,\mathcal{E}[W[n]]\Big)\\&=\sum_{m=1}^{k_1}[p+p^2(k_2-1)]=pk_1[1+p(k_2-1)]\end{aligned} \tag{6-167}$$

自协方差函数可由自相关函数减去$\mathcal{E}[X[k_1]]\mathcal{E}[X[k_2]]=p^2k_1k_2$ 得到：

$$C_{XX}[k_1,k_2]=pk_1[1+p(k_2-1)]-p^2k_1k_2=pqk_1 \tag{6-168}$$

这样，对一般的 $\{k_1,k_2\}$ 有

$$R_{XX}[k_1,k_2]=pq\min(k_1,k_2)+p^2k_1k_2 \tag{6-169}$$

$$C_{XX}[k_1,k_2]=pq\min(k_1,k_2) \tag{6-170}$$

从上式中我们看到 $X[k]$不是广义平稳的。因为 k 时刻的方差为 $\mathrm{var}[X[k]]=pqk$，相关系

数为：

$$\rho=\frac{pq\min(k_1,k_2)}{\sqrt{(pqk_1)(pqk_2)}}=\frac{\min(k_1,k_2)}{\sqrt{k_1k_2}} \tag{6-171}$$

表 6-4 中总结了以上特性，还包括伯努利游走以及 6.12 节讨论的随机过程的一些结果。有趣的是一些大家熟知的随机过程的相关系数具有相同的形式，而且独立于其参数。

表 6-4 随机过程的相关和协方差

过程名	自相关函数	自协方差函数	相关系数
伯努利游走	$4pq\min(k_1, k_2)+(p-q)^2k_1k_2$	$4pq\min(k_1, k_2)$	$\min(k_1, k_2)/\sqrt{k_1k_2}$
二项式计数	$pq\min(k_1, k_2)+p^2k_1k_2$	$pq\min(k_1, k_2)$	$\min(k_1, k_2)/\sqrt{k_1k_2}$
泊松计数	$\lambda\min(t_1, t_2)+\lambda^2t_1t_2$	$\lambda\min(t_1, t_2)$	$\min(t_1, t_2)/\sqrt{t_1t_2}$
"半随机"电报	$\exp(-2\lambda\|t_2-t_1\|)$	$\exp(-2\lambda\|t_2-t_1\|)-\exp(-2\lambda(t_1+t_2))$	see(6-206)
"随机"电报	$\exp(-2\lambda\|t_2-t_1\|)$	$\exp(-2\lambda\|t_2-t_1\|)$	$\exp(-2\lambda\|t_2-t_1\|)$
维纳	$\alpha\varepsilon^2\min(t_1, t_2)$	$\alpha\varepsilon^2\min(t_1, t_2)$	$\min(t_1, t_2)/\sqrt{t_1t_2}$

6.12 随机过程

接下来，我们考虑一些随着时间不断发展的随机过程。对于随机序列，由时间索引的随机变量可以具有离散的或连续的输出。我们先从具有泊松分布的独立随机变量序列开始，然后允许时间间距趋于 0，以产生一个连续时间的随机过程，这样可以在任何时刻在振幅上实现离散变化。(注意，各种随机过程的曲线的例子实际上是离散时间的现实，因为 MATLAB 只能绘制过程的有限个输出。)

6.12.1 泊松计数过程

泊松随机变量的输出是离散的(无限可数)，由所有非负整数给定。定义时间间隔$\Delta t\triangleq t/k$，其中 $t\in\mathcal{T}=\mathcal{R}^+$为连续时间，$k\in\mathcal{N}$。考虑随机过程，其定义如下：

$$X(t)\triangleq X[k\Delta t] \tag{6-172}$$

其中 $X[k](\Delta t=1)$为式(6-160)中二项式计数序列。如果等号右边的变量扩展，伯努利成功就可在任意期间间隔发生(而不仅仅是 $\Delta t=1$)。我们假设以下条件，当 $\Delta t\to 0$ 时，$X(t)$ 是泊松计数过程(对所有 $t\in\mathcal{T}$)：

- $X[k\Delta t]$为一个二项随机变量，其成功概率 p 与时间间隔成正比：

$$p=P(X[k\Delta t]=1)=\lambda\Delta t \tag{6-173}$$

  其中 λ 为比例参数。

- 间隔 Δt 足够小，因此在任意间距内最多有一次伯努利成功。在任意间距有两次或更多成功的概率假设为 0。

对二项式计数过程和以上的条件，n 次成功的概率为：

$$P(X[k\Delta t]=n)=\binom{k}{n}(\lambda\Delta t)^n(1-\lambda\Delta t)^{k-n},n=0,\cdots,k \tag{6-174}$$

第 4 章中提到过这种二项式分布，它可用泊松分布来近似。特别是，如果$k\to\infty$，则间隔 $\Delta t\to 0$，因此 $p\to 0$，满足约束条件 $kp=\lambda t<\infty$为常数，则式(6-174)可做如下近似：

$$P(X(t)=n)\approx\frac{(\lambda t)^n}{n!}\exp(-\lambda t),n\in\mathcal{Z}^+ \tag{6-175}$$

上式服从泊松分布，时间依赖参数为 $\alpha=\lambda t$。

定义(泊松计数过程) **泊松计数过程** $X(t)$是一个离散输出，连续时间随机过程具有 $X(0)=0$，因此在 t 时刻输出的概率由式(6-175)中参数为 $\alpha=\lambda t$ 的泊松分布给出。

与二项式计数序列类似，这个泊松过程的现实只有正的单位振幅的步进。但是，不同于二项式计数序列的是，二项式计数序列的输出只在离散的时间点上发生，而泊松计数过程的振幅可在任意时间 $t\in\mathcal{T}$ 改变。图 6-19 中给出了它的一个现实示例。此泊松过程的均值和方差可以很容易地由(6-175)得到：$\mathcal{E}[X(t)]=\operatorname{var}[X(t)]=\lambda t$。由于这些矩都与时间相关，所以过程非平稳。

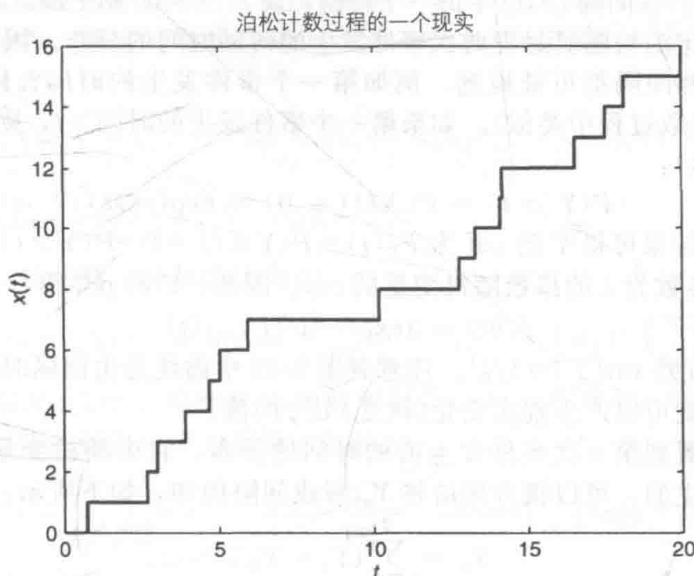


图 6-19 比例参数为 $\lambda=1$ 的泊松计数过程的现实示例

例 6-24 泊松过程 $X(t)$是许多物理过程的一个精确模型，比如放射衰减以及电话线故障问题。假设客户以每小时 $\lambda=5$ 的速率进入一个商场，则各种概率很容易确定，比如：(a)在三小时内来 8 个顾客的概率；(b)在前一个小时没有顾客到达的条件下，3 个顾客在一个小时内到达的概率；(c)一个小时内有 1 个顾客到达，接下来一个小时 2 个顾客到达的概率。相应的概率描述是：(a)$P(X(3)=8)$；(b)$P(X(2)-X(1)=3\mid X(1)-X(0)=0)$；(c)$P(X(1)=1,\ X(2)=3)$。它们都对应于初始条件 $X(0)=0$。(a)可以直接得出：

$$P(X(3)=8)=\frac{(5\times 3)^8}{8!}\exp(-5\times 3)\approx 0.0194 \tag{6-176}$$

(b)中，条件可以不考虑，因为增量是独立的：

$$P(X(2)-X(1)=3)=\frac{(5\times 1)^3}{3!}\exp(-5\times 1)\approx 0.0056 \tag{6-177}$$

(c)可通过将表达式写为独立增量的形式来处理：

$P(X(1)=1,\ X(2)=3)=P(X(1)=1,\ X(2)-X(1)=2)=P(X(1)=1)P(X(2)-X(1)=2)$，因此：

$$P(X(1)=1,X(2)=3)=\frac{(5\times 1)^1}{1!}\exp(-5\times 1)\,\frac{(5\times 1)^2}{2!}\exp(-5\times 1)\approx 0.0028 \tag{6-178}$$

泊松过程是莱维随机过程的一种特殊类型。◀

定义(莱维随机过程) **莱维随机过程** $X(t)$是具有以下特性的任意连续时间随机过程：(ⅰ)isi；(ⅱ)零初始输出 $X(0)=0$(几乎是确定的)；(ⅲ)右连续(几乎必然)，并具有左极限。

回忆右连续函数 $x(t)$，满足对 $\xi>0$ 和 $t\in\mathcal{T}$，有$\lim\limits_{\xi\to 0}x(t+\xi)=x(t)$。如果$\lim\limits_{\xi\to 0}x(t-\xi)$存在，则在 t 时刻，它有左极限。离散随机变量的 cdf 是一个具有左极限的右连续函数的一个例子。

其他莱维随机过程的例子还有伽马和维纳过程，将在稍后讨论。由于泊松过程是 isi，在非重叠间隔上发生的事件是独立的，它们的分布仅依赖于间隔的持续时间，而不依赖于它们在$\mathcal{T}$的位置。这意味着对于持续时间 $\tau>0$，有

$$P(X(t)-X(t-\tau)=n)=\frac{(\lambda\tau)^n}{n!}\exp(-\lambda\tau) \tag{6-179}$$

式(6-175)是式(6-179)在间隔 $[0,t]$ 中的一个特例，设 $t-\tau=0$，则 $\tau=t$，$X(t-\tau)\to X(0)=0$。

我们还可以确定泊松随机过程两次事件发生的间隔时间的分布。因为泊松过程是一个 isi 过程，$\mathcal{T}$中的任意间隔都可被检测，例如第一个事件发生的时间，用随机变量 Y 表示(与之前在二项式计数过程中类似)。如果第一个事件发生的时间$>t$，则 $[0,t]$ 间不会再有事件发生(反之亦然)：

$$P(Y>t)=P(X(t)=0)=\exp(-\lambda t) \tag{6-180}$$

其中 $t\geqslant 0$。由这个结果可得 Y 的 cdf 为 $F_Y(t)=P(Y\leqslant t)=1-P(Y>t)=[1-\exp(-\lambda t)]I_{[0,\infty)}(t)$，这就是参数为 λ 的指数随机变量的 cdf。因此，Y 的 pdf 为：

$$f_Y(t)=\lambda\exp(-\lambda t)I_{[0,\infty)}(t) \tag{6-181}$$

均值$\mathcal{E}[Y]=1/\lambda$，方差 $\mathrm{var}[Y]=1/\lambda^2$。注意到图 6-19 中曲线是由间隔时间满足指数分布的样本产生的，这样就可以产生现实变化(跳变)处 t 的值。

这个结果可扩展到第 n 次事件发生前的时间的分布，它由随机变量 Y_n 定义。由于事件发生的间隔是独立的，可以很方便的将 Y_n 写成间隔的和，如下所示：

$$Y_n=\sum_{k=1}^{n}(Y_k-Y_{k-1}) \tag{6-182}$$

因为 Y_n 是独立的指数随机变量的和，我们可由第 4 章的例 4-17 推论得到 Y_n 是一个埃尔朗(伽马)随机变量，参数为$\{\lambda,\ r=n\}$：

$$f_{Y_n}(t)=\frac{\lambda^n t^{n-1}}{(n-1)!}\exp(-\lambda t)I_{[0,\infty)}(t) \tag{6-183}$$

均值$\mathcal{E}[Y_n]=n/\lambda$，方差 $\mathrm{var}[Y_n]=n/\lambda^2$。如我们预料的，均值和方差随着事件发生的数目 n 增长。

接下来，我们分析在 $[0,t]$ 内等间隔单次到达的条件下，首次到达时间 Y 的分布。由于 $X(t)$是记录 $[0,t]$ 间到达数目的泊松随机变量，由条件概率可得：

$$\begin{aligned}P(Y\leqslant y|X(t)=1)&=P(Y\leqslant y,X(t)=1)/P(X(t)=1)\\&=P(X(y)=1,X(t)=1)/P(X(t)=1)\end{aligned} \tag{6-184}$$

其中用到了 $Y\leqslant y\Rightarrow X(y)=1$。由于：

$$P(X(y)=1,X(t)=1)=P(X(y)=1,X(t)-X(y)=0) \tag{6-185}$$

并且泊松过程具有独立增量，式(6-184)变为：

$$\begin{aligned}P(Y\leqslant y|X(t)=1)&=P(X(y)=1)P(X(t)-X(y)=0)/P(X(t)=1)\\&=\frac{\lambda y\exp(-\lambda y)\exp(-\lambda(t-y))}{\lambda t\exp(-\lambda t)}=(y/t)I_{[0,t]}(y)\end{aligned} \tag{6-186}$$

其中用指示函数来强调 $y\in[0,t]$。这个结果是一个均匀分布随机变量的 cdf。它说明在 $[0,t]$ 间且 $X(t)=n$ 个到达的条件下，到达次数的条件分布是：在 $[0,t]$ 上均匀分布的随机变量的 n 个 iid 样本的顺序统计量的分布(参见习题 6-37)。

可以看出泊松计数过程可以建模为滤波后狄拉克 δ 函数链。该滤波器的冲激响应是连续单位阶跃函数 $u(t)$，冲激之间的间隔具有指数分布。由单位阶跃函数滤波的信号等价于一个积分器，它是之前讨论的伯努利游走和二项式计数过程的求和滤波器的连续时间版本。这并不奇怪，因为在引出泊松计数过程时，我们从二项式计数过程开始，让时间间隔 $\Delta t\to 0$。二项式计数过程的滤波器的输入包括克罗内克 δ 函数，此函数相邻时间间隔具有几何分布(见式(6-164))。实际上，该滤波器模型的离散的单位阶跃函数在 $\Delta t\to 0$ 时被泊

松计数过程的连续单位阶跃函数所取代。狄拉克 δ 函数之间的非均匀间距的原因来自二项式计数过程的零字符串，二项式计数过程也是由此导出。

为了计算 $X(t)$ 的自相关函数，需要泊松过程在两个时刻 $\{t_1, t_2\}$ 的联合概率分布。这可以通过运用如下 $t_2>t_1$ 时的条件概率得到：

$$P(X(t_1)=n_1, X(t_2)=n_2)=P(X(t_2)=n_2 \mid X(t_1)=n_1)P(X(t_1)=n_1) \quad (6\text{-}187)$$

其中 $n_2>n_1$，由于指数分布的无记忆特性(见第 3 章)，条件概率呈参数为 $\{\lambda(t_2-t_1), r=n_2-n_1\}$ 的埃尔朗分布。假定 t_1 前发生的 n_1，分布仅由时间间隔 $(t_1, t_2]$ 内额外发生的 n_2-n_1 的数目决定。因此：

$$\begin{aligned}P(X(t_1)=n_1, X(t_2)=n_2)&=\frac{[\lambda(t_2-t_1)]^{n_2-n_1}}{(n_2-n_1)!}\exp(-\lambda(t_2-t_1))\frac{(\lambda t_1)^{n_1}}{n_1!}\exp(-\lambda t_1)\\&=\frac{[\lambda(t_2-t_1)]^{n_2-n_1}(\lambda t_1)^{n_1}}{(n_2-n_1)!n_1!}\exp(-\lambda t_2)\end{aligned} \quad (6\text{-}188)$$

则此指数项仅为 t_2 的函数，将其简化为：

$$P(X(t_1)=n_1, X(t_2)=n_2)=\frac{1}{n_2!}[\lambda(t_2-t_1)]^{n_2-n_1}(\lambda t_1)^{n_1}\binom{n_2}{n_1}\exp(-\lambda t_2) \quad (6\text{-}189)$$

对于 $n_2>n_1$ 以及 $t_2>t_1$，自相关函数的期望是一个二重求和，因为式(6-189)是一个离散的联合 pdf(即使过程是定义为连续时间的)：

$$\begin{aligned}R_{XX}(t_1, t_2)&=\sum_{n_1=0}^{n_2}\sum_{n_2=0}^{\infty}n_1 n_2\frac{[\lambda(t_2-t_1)]^{n_2-n_1}(\lambda t_1)^{n_1}}{n_2!}\binom{n_2}{n_1}\exp(-\lambda t_2)\\&=\sum_{n_2=0}^{\infty}\frac{n_2}{n_2!}\exp(-\lambda t_2)(\lambda t_2)^{n_2}\sum_{n_1=0}^{n_2}n_1\binom{n_2}{n_1}(t_1/t_2)^{n_1}[1-t_1/t_2]^{n_2-n_1}\end{aligned} \quad (6\text{-}190)$$

由于第二个求和给出了一个参数为 $\{p=t_1/t_2, N=n_2\}$ 的二项式分布的均值 Np，我们得到：

$$R_{XX}(t_1, t_2)=\sum_{n_2=0}^{\infty}\frac{n_2}{n_2!}\exp(-\lambda t_2)(\lambda t_2)^{n_2}n_2(t_1/t_2)=(t_1/t_2)\sum_{n_2=0}^{\infty}n_2^2\frac{(\lambda t_2)^{n_2}}{n_2!}\exp(-\lambda t_2) \quad (6\text{-}191)$$

它是参数为 $\alpha=\lambda t_2$ 的泊松分布的二阶矩 $\alpha(1+\alpha)$，则：

$$R_{XX}(t_1, t_2)=(t_1/t_2)\lambda t_2(1+\lambda t_2)=\lambda t_1(1+\lambda t_2) \quad (6\text{-}192)$$

其中 $t_2>t_1$。最终，对于任意 $\{t_1, t_2\}$：

$$R_{XX}(t_1, t_2)=\lambda\min(t_1, t_2)+\lambda^2 t_1 t_2 \quad (6\text{-}193)$$

它不依赖于时间间隔 $\tau=t_2-t_1$，因此 $X(t)$ 不是广义平稳的。图 6-20a 为 $R_{XX}(t_1, t_2)$ 的曲线，原点处有个马鞍状的点。由于 $\mathcal{E}[X(t)]=\lambda t$，则自协方差函数为：

$$C_{XX}(t_1, t_2)=R_{XX}(t_1, t_2)-\mathcal{E}[X(t_1)]\mathcal{E}[X(t_2)]=\lambda\min(t_1, t_2) \quad (6\text{-}194)$$

表 6-4 也总结了泊松过程的这些结果。表 6-4 中我们还发现相关系数与二项式过程的相关系数具有相似的形式(除了它是对连续时间定义的)。图 6-20b 中画出了相关系数的示意图。

λ 为常数的这类泊松过程被称为齐次的(或一致的)。速率参数可扩展为一个时间函数 $\lambda(t)$，对应于事件的速率随时间变化的情况(比如一天的不同时间)。这种非齐次(或非一致)泊松过程的分布为：

$$P(X(t)=n)=\frac{1}{n!}\left(\int_0^t\lambda(v)\mathrm{d}v\right)^n\exp\left(-\int_0^t\lambda(v)\mathrm{d}v\right) \quad (6\text{-}195)$$

$X(t)$ 的均值和方差又与 t 相关：$\mathcal{E}[X(t)]=\mathrm{var}[X(t)]=\int_0^t\lambda(v)\mathrm{d}v$，对齐次的情况则为 λt。$X(t)$ 的增量是独立的，但对于齐次过程的情况是非平稳的。与式(6-179)类似，我们可通过改变积分限将任意间隔 τ 的输出的概率写为：

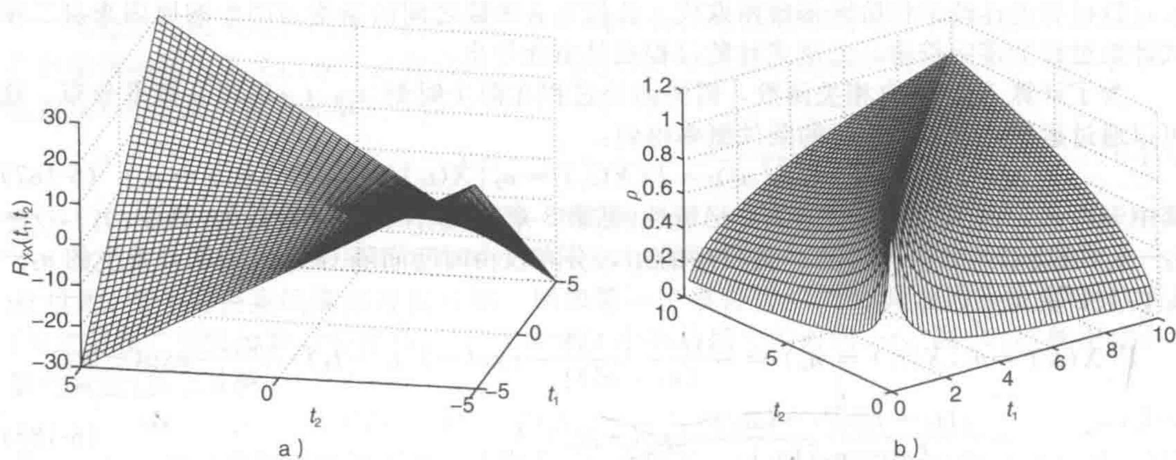


图 6-20 $\lambda=1$ 的泊松计数过程的相关性。a)泊松计数过程的自相关函数 $R_{XX}(t_1,t_2)$；b)泊松计数过程的相关系数 ρ。这也是维纳过程、伯努利游走和二项式计数过程(但是对于 $\{t_1,t_2\}$ 内的整数值)的相关系数的示意图

$$P(X(t)-X(t-\tau)=n)=\frac{1}{n!}\left(\int_{t-\tau}^{t}\lambda(v)\mathrm{d}v\right)^n\exp\left(-\int_{t-\tau}^{t}\lambda(v)\mathrm{d}v\right) \tag{6-196}$$

非齐次情况的自相关函数可由式(6-193)得到，只需用 $\int_0^{\min(t_1,t_2)}\lambda(v)\mathrm{d}v$ 代替 $\min(t_1,t_2)$ 即可。

最后，我们将泊松计数过程扩展到可用伽马分布来描述到达时间(而不是指数分布)：

$$f_Y(y)=\frac{y^{r-1}\lambda^r}{\Gamma(r)}\exp(-\lambda y)I_{[0,\infty)}(y) \tag{6-197}$$

参数为 $\{\lambda,r>0\}$。

定义(伽马随机过程) **伽马随机过程**是一个具有伽马到达时距的莱维随机过程。

图 6-21 给出了现实的示例，从中我们可以看到计数速率随着 r 的增大而减小，因为伽马分布的均值随着 r 增大而增大。图 6-19 中的泊松现实处于显示的两种情况之间，因为指数到达间隔分布对应于 $r=1$。

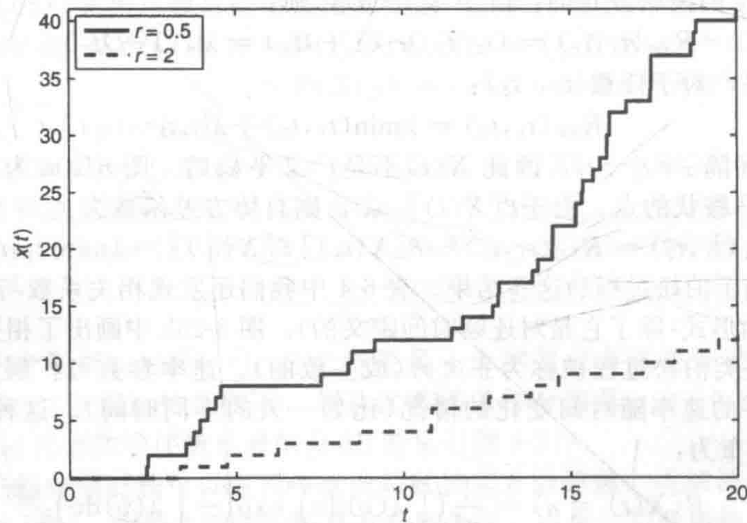


图 6-21 对应不同的 r 值，速率参数 $\lambda=1$ 的伽马过程的现实示例

6.12.2 随机电报信号

定义(随机电报信号) **随机电报信号**可由如下的泊松计数过程 $X(t)$ 得到：

$$Y(t)=\begin{cases}1, & X(t)\text{ 是偶数}\\ -1, & X(t)\text{ 是奇数}\end{cases} \tag{6-198}$$

可以简写为 $Y(t)=(-1)^{X(t)}$，其中 $X(t)$的分布如式(6-175)所示，在$\mathcal{T}=[0, t]$上。

不同于记录步进的数目，$Y(t)$仅标示了计数值是奇数或偶数。由于 $X(0)=0$，则随机电报信号的初始条件为 $Y(0)=1$。$Y(t)$是一个连续时间过程，一直保持当前值(1 或−1)直到泊松过程 $X(t)$发生变化。

尽管会出现输出是等可能的情况，但实际上只有当 $t\to\infty$时才为真。对于有限 t，泊松过程 $X(t)$为偶数次步进的概率为：

$$P(X(t)\text{ 是偶数})=\sum_{m=0,\text{偶数}}^{\infty}\exp(-\lambda t)(\lambda t)^m/m! \tag{6-199}$$

其中用到了式(6-175)中计数的泊松分布公式。提出 $\exp(-\lambda t)$因子，我们看到无限级数是双曲余弦函数(见附录 E)：

$$P(X(t)\text{ 是偶数})=\exp(-\lambda t)\cosh(\lambda t) \tag{6-200}$$

对 $X(t)$为奇数情况有：

$$P(X(t)\text{ 是奇数})=\exp(-\lambda t)\sinh(\lambda t) \tag{6-201}$$

综合以上结果得到 $Y(t)$的 pdf：

$$f_{Y(t)}(y)=\exp(-\lambda t)[\cosh(\lambda t)\delta(y-1)+\sinh(\lambda t)\delta(y+1)] \tag{6-202}$$

其中 $\delta(y)$为狄拉克 δ 函数。注意到 $f_{Y(t)}(y)$是一个有效的 pdf：

$$\int_{-\infty}^{\infty} f_{Y(t)}(y)\mathrm{d}y=\exp(-\lambda t)[\cosh(\lambda t)+\sinh(\lambda t)]=1 \tag{6-203}$$

图 6-22 是 $\lambda=1$ 的曲线图，从图中可以看出随着 t 的增大，式(6-202)中用来对 $\delta(y-1)$和 $\delta(y+1)$加权的分量分别衰减，但在任意时间 t，它们的和一直为 1。只有在 $t\to\infty$时，它们相等(都为 1/2)。实际情况是，3 到 4 个时间常数后它们就几乎相等了，其中指数函数的时间常数为 $\tau\approx 1/\lambda$。

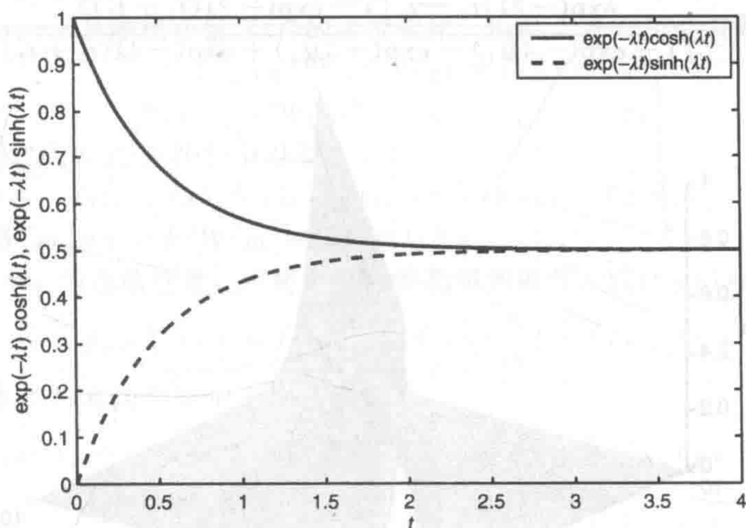


图 6-22 $\lambda=1$ 时的随机电报信号的 pdf 分量

从滤波的角度看，一个时不变滤波器作用于一个狄拉克 δ 函数链是不可能产生随机电报信号的。回想一下，对于泊松计数过程这是可能的，其中所述滤波器具有由单位阶跃函数给定的冲激响应。此处的滤波器是矩形的，时变期间由 δ 函数的间隔来确定。

随机电报过程的均值为：

$$\mathcal{E}[Y(t)]=\exp(-\lambda t)[(+1)\cosh(\lambda t)+(-1)\sinh(\lambda t)]=\exp(-2\lambda t) \tag{6-204}$$

由于二阶矩为$\mathcal{E}[Y^2(t)]=1$，方差为 $\mathrm{var}[Y(t)]=1-\exp(-4\lambda t)$，随时间变化。但是随着

$t\to\infty$，随机电报信号具有零均值和单位方差，与对称的伯努利序列相似。图 6-23 给出的是 $\lambda=1$ 的随机电报信号的现实示例。随机化的随机电报信号由 $Z(t)=WY(t)$ 给定。其中 W 是一个具有等可能输出 $\{-1,1\}$ 的对称伯努利随机变量，而且与 $Y(t)$ 独立。（随机电报信号的初始版本也被称为"半随机信号"）。加入这个附加的随机性后初始条件变为 $Y(0)=W$，我们不再知道是过程的 1 还是 -1 与跳变的偶数次对应。

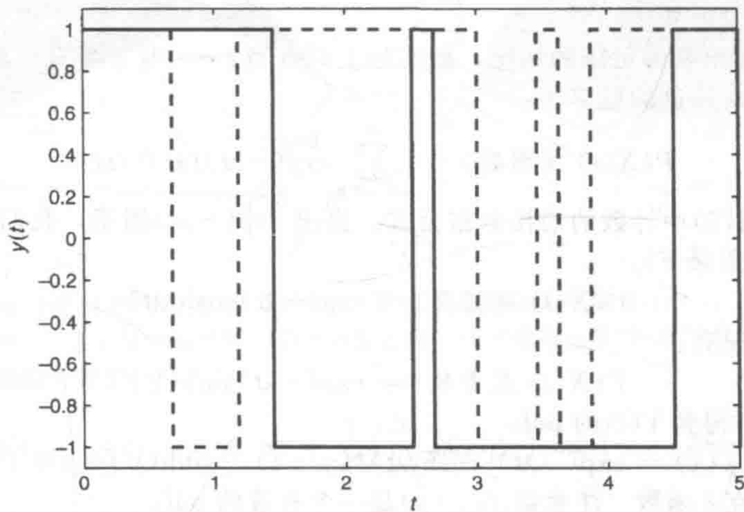


图 6-23 $\lambda=1$ 时的随机电报信号的现实示例

可以证明随机电报信号的自协方差函数为（见习题 6-31）

$$C_{ZZ}(t_1,t_2)=\exp(-2\lambda|t_2-t_1|)=C_{ZZ}(\tau) \tag{6-205}$$

它是广义平稳的。$C_{ZZ}(t_1,t_2)$ 也是相关系数 ρ，因为对于所有的 t，$Z(t)$ 的方差为 1。图 6-24画了相关系数的示意图。$Y(0)=1$ 的"半随机"电报信号的相关系数为（见题 6-32）

$$\rho=\frac{\exp(-2\lambda|t_2-t_1|)-\exp(-2\lambda(t_1+t_2))}{\sqrt{1-\exp(-4\lambda t_1)-\exp(-4\lambda t_2)+\exp(-4\lambda(t_1+t_2))}} \tag{6-206}$$

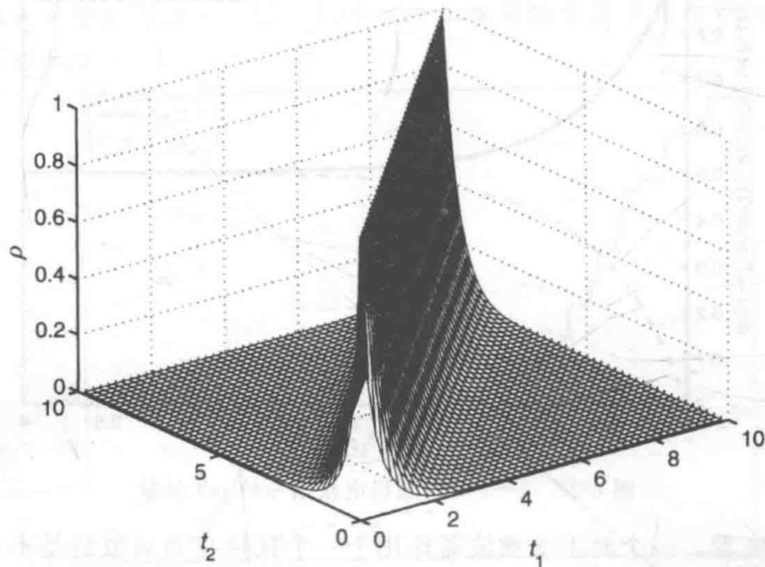


图 6-24 "随机化"的电报信号（$\lambda=1$）的相关系数。"半随机"电报信号（$Y(0)=1$）的相关系数与此图相似，除了较小的 $\{t_1,t_2\}$

它已不能简化了。但注意到对于大的 $\{t_1,t_2\}$，两种版本的随机电报信号的相关系数近似相同。表 6-4 中总结了两种随机电报信号的相关结果。

6.12.3 维纳过程

维纳过程是布朗运动这种物理现象的一个模型。布朗运动是由于水分子频繁的随机碰撞而形成的颗粒的不规律运动。这种模型可从对称伯努利序列$W[k]$开始推导，但对于小的$\varepsilon>0$有输出$\{-\varepsilon,\varepsilon\}$。假设颗粒的碰撞是等时间间距以速率$\alpha$发生，也就是说它们在$1/\alpha$的整数倍发生。到时间$t$时总的碰撞次数$N=\lfloor \alpha t \rfloor$，其中$\lfloor \alpha \rfloor$是保持最大整数$\leqslant\alpha$的底函数。图 6-25 给出了各种量值。定义随机过程

$$X(t)=\sum_{m=1}^{\lfloor \alpha t \rfloor}W[m] \tag{6-207}$$

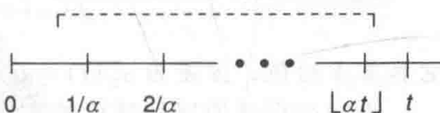


图 6-25 导出维纳过程$X(t)$用到的离散时间间隔

它与$p=1/2$的伯努利随机游走类似(见表 6-3)；这里的差别是：(ⅰ)离散时间间隔持续期为$1/\alpha$而不是1；(ⅱ)振幅是$\pm\varepsilon$而不是±1；(ⅲ)求和的上限是$\lfloor \alpha t \rfloor$而不是k。因为这个上限，$X(t)$会一直保持当前值直到$t\to t+1/\alpha$，此时因为另一次碰撞使得求和式中包括一个新项。这是扩展对称伯努利序列为一个连续时间t的函数的第一步。就如对泊松计数过程一样，维纳过程可通过令$1/\alpha\to0$来得到，最终得到一个连续时间过程。而且，振幅$\varepsilon\to0$使得在每个时刻的随机变量是连续的，且具有明确的 pdf。由于：

$$\mathcal{E}[W[m]]=0,\quad \mathcal{E}[W^2[m]]=\varepsilon^2,\quad 则\mathcal{E}[X(t)]=0,\quad \mathcal{E}[X^2(t)]=\varepsilon^2\lfloor \alpha t \rfloor$$

假设当前时间t与$1/\alpha$的整数倍相近，$f_{X(t)}(x)$存在。考虑下一个时刻$t+1/\alpha$的 pdf，应用全概率公式得到：

$$\begin{aligned}f_{X(t+1/\alpha)}(x)=&f_{X(t+1/\alpha)}(x\,|\,W[m]=\varepsilon)P(W[m]=\varepsilon)\\&+f_{X(t+1/\alpha)}(x\,|\,W[m]=-\varepsilon)P(W[m]=-\varepsilon)\end{aligned} \tag{6-208}$$

其中假设$m=\lfloor \alpha t \rfloor$正好对应$t$时刻。这样分割描述了过程从时间$t$到$t+1/\alpha$时可能会发生的两个事件。为了重新用$t$时刻的 pdf 来描述式(6-208)，我们设定$X(t+1/\alpha)=X(t)+\varepsilon$，$X(t+1/\alpha)=X(t)-\varepsilon$是随机变量$X(t)$的线性变化。由第 4 章的变换结果：

$$f_{X(t+1/\alpha)}(x\,|\,W(m]=\varepsilon)=f_{X(t)}(x-\varepsilon) \tag{6-209}$$

$$f_{X(t+1/\alpha)}(x\,|\,W[m]=-\varepsilon)=f_{X(t)}(x+\varepsilon) \tag{6-210}$$

将这些表达式代入式(6-208)中得到：

$$f_{X(t+1/\alpha)}(x)=(1/2)[f_{X(t)}(x-\varepsilon)+f_{X(t)}(x+\varepsilon)] \tag{6-211}$$

其中用到了$P(W[m]=\varepsilon)=P(W[m]=-\varepsilon)=1/2$。

由于假设t和ε为连续变量，可将函数用泰勒级数展开。式(6-211)的左边对t展开为：

$$f_{X(t+1/\alpha)}(x)=f_{X(t)}(x)+(1/\alpha)\frac{\partial}{\partial t}f_{X(t)}(x)+\cdots \tag{6-212}$$

同样，式(6-211)右边对x展开为：

$$f_{X(t)}(x+\varepsilon)=f_{X(t)}(x)+\varepsilon\frac{\partial}{\partial x}f_{X(t)}(x)+\varepsilon^2\frac{\partial^2}{\partial x^2}f_{X(t)}(x)+\cdots \tag{6-213}$$

$$f_{X(t)}(x-\varepsilon)=f_{X(t)}(x)-\varepsilon\frac{\partial}{\partial x}f_{X(t)}(x)+\varepsilon^2\frac{\partial^2}{\partial x^2}f_{X(t)}(x)-\cdots \tag{6-214}$$

这些式子代表了$f_{X(t+1/\alpha)}(x)$沿时间轴t和振幅轴x展开的形式。由于最终目的是让$\varepsilon\to0$(创造一个连续振幅函数)以及$\alpha\to\infty$(允许过程在任意时间变化)，我们忽略所有高阶项，因为它们要么依赖$1/\alpha^m(m\geqslant2)$要么是$\varepsilon^m(m\geqslant3)$。注意到我们必须保留$\varepsilon^2$项，因为：(ⅰ)含$\varepsilon$的两项互相抵消；(ⅱ)$\alpha\varepsilon^2$假定为常数。

将式(6-212)～式(6-214)代入式(6-211)得到：

$$(1/\alpha)\frac{\partial}{\partial t}f_{X(t)}(x)=\varepsilon^2\frac{\partial^2}{\partial x^2}f_{X(t)}(x)+\cdots \tag{6-215}$$

其中省略号表示的是两个展开式的高阶项。最后，令 $\alpha\to\infty$，$\varepsilon\to 0$，则 $\alpha\varepsilon^2$ 为常数，舍弃所有高阶项，得到二阶偏微分方程(DE)：

$$\frac{\partial}{\partial t}f_{X(t)}(x)=\alpha\varepsilon^2\frac{\partial^2}{\partial x^2}f_{X(t)}(x) \tag{6-216}$$

这个方程的解是维纳过程的 pdf(见题 6-33)。

定义(维纳过程) **维纳过程**是一个连续时间过程，具有如下高斯 pdf：

$$f_{X(t)}(x)=\frac{1}{\sqrt{2\pi\alpha\varepsilon^2 t}}\exp(-x^2/2\alpha\varepsilon^2 t) \tag{6-217}$$

它具有零均值，方差为 $\sigma_X^2(t)=\alpha\varepsilon^2 t$。初始条件是 $X(0)=0$。

维纳过程是伯努利随机游走的时间连续、振幅连续的情况，它的振幅可在任意时刻 $t\in\mathcal{T}=\mathcal{R}^+$ 变化，且振幅可以为 $\mathcal{R}$ 中任意值。因此，与泊松计数过程相似，维纳过程可被看作是一个积分滤波器的输出。差别在于模型中设定伯努利序列的输出为 $\pm\varepsilon$(而不是泊松计数过程的 $\{0,1\}$)。第 8 章中将证明，一个维纳过程可通过对 PSD 为 $\alpha\varepsilon^2$ 的白噪声积分产生。尽管维纳过程的均值始终为零，但它的方差 $\alpha\varepsilon^2 t$ 随时间线性增长，因此 $X(t)$ 是一个非平稳随机过程。图 6-26 是维纳过程的现实示例。

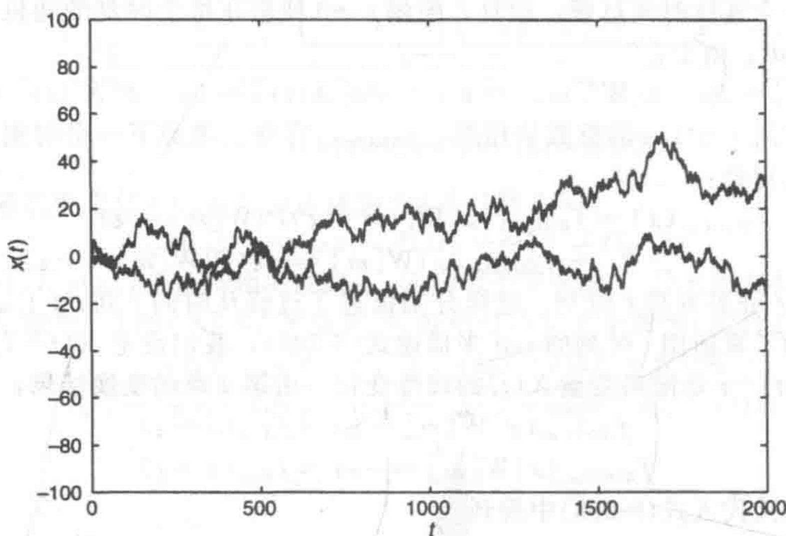


图 6-26 式(6-217)中 $\alpha\varepsilon^2=1$ 时维纳过程的现实示例

维纳过程可用更通用的形式定义，假设它是一个 isi 过程(因此是莱维过程的一类)

定义(扩展的维纳过程) **扩展的维纳过程** $X(t)$ 是一个 isi 过程，初始值 $X(0)=0$，因此有：

$$f_{X(t_2)-X(t_1)}(x)=\frac{1}{\sqrt{2\pi\alpha\varepsilon^2(t_2-t_1)}}\exp(-x^2/2\alpha\varepsilon^2(t_2-t_1)) \tag{6-218}$$

其中 $t_2>t_1$。

$X(t_2)-X(t_1)$ 的方差为 $\alpha\varepsilon^2(t_2-t_1)$，在式(6-217)中当 $t_2=t$，$t_1=0$ 时，方差变为 $\alpha\varepsilon^2 t$。$X(t_2)$ 和 $X(t_1)$ 的自协方差函数为：

$$\begin{aligned}C_{XX}(t_1,t_2)&=\mathcal{E}[X(t_2)X(t_1)]=R_{XX}(t_1,t_2)=\mathcal{E}[(X(t_2)-X(t_1)+X(t_1))X(t_1)]\\&=\mathcal{E}[(X(t_2)-X(t_1))X(t_1)]+\mathcal{E}[X^2(t_1)]\end{aligned} \tag{6-219}$$

由于增量是独立的，最后一行的第一个期望值分成两个期望值的乘积，则

$$\begin{aligned}C_{XX}(t_1,t_2)&=\mathcal{E}[X(t_2)-X(t_1)]\,\mathcal{E}[X(t_1)]+\mathcal{E}[X^2(t_1)]\\&=\mathcal{E}[X^2(t_1)]=\alpha\varepsilon^2 t_1\end{aligned} \tag{6-220}$$

其中我们用到了过程具有零均值的特性来消去式子的第一项。这个结果是在假设 $t_2>t_1$ 时得到的，对于任意 $\{t_1,t_2\}$：

$$C_{XX}(t_1,t_2)=\alpha\varepsilon^2\min(t_1,t_2)=R_{XX}(t_1,t_2) \tag{6-221}$$

这也是自相关函数，因为过程是零均值的，图 6-27 画出了它的示意图。注意到泊松计数过程的自协方差函数具有与这张图相同的形状，同样的还有伯努利游走和泊松计数过程(但是在 $\{t_1,t_2\}$ 的整数值)。维纳过程的相关性的结果也在表 6-4 中进行了总结。

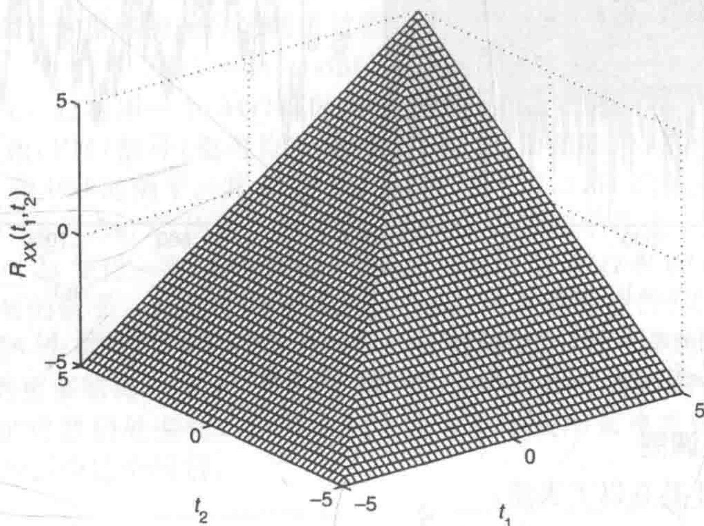


图 6-27 $\alpha\varepsilon^2=1$ 的扩展维纳过程的自相关函数

6.12.4 高斯过程

考虑一个随机过程，其输出由高斯分布给定。

定义(高斯过程) **高斯过程** $X(t)$是一组由连续时间索引的随机变量，则对任意 N 个时刻 $\{t_1,\cdots,t_N\}$，它们的联合 pdf 是高斯的：

$$f_{X(t_1),\cdots,X(t_N)}(x_1,\cdots,x_N)=\frac{1}{\sqrt{(2\pi)^N|\boldsymbol{C}_{XX}|}}\exp(-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}_{XX}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \tag{6-222}$$

其中 $\boldsymbol{X}\triangleq[X(t_1)\cdots,\ X(t_N)]^{\mathrm{T}}$，$\boldsymbol{x}\triangleq[x_1,\ \cdots x_N]^{\mathrm{T}}$，$\boldsymbol{\mu}=\mathcal{E}[\boldsymbol{X}]$，$\boldsymbol{C}_{XX}=\mathcal{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}]$ 是自协方差矩阵，$|\boldsymbol{C}_{XX}|$是 $\boldsymbol{C}_{XX}$ 的行列式。

图 6-28a 是一个 iid 高斯过程的一个现实示例，参数 $\boldsymbol{\mu}=\boldsymbol{0}$，$\boldsymbol{C}_{XX}=\boldsymbol{I}$。由于之前所有的过程都是基于对伯努利序列的求和或积分得到的，自相关函数由过程如何构造决定。而高斯过程却不是以这样一种模式定义的：协方差矩阵是过程的一个独立参量，由式(6-222)中 $\boldsymbol{C}_{XX}$ 给定。与高斯随机向量类似，一个高斯随机过程的特性完全由其均值向量和协方差矩阵决定(可以是时间的函数)。

假设我们要产生相关高斯过程的一个现实。设 $\boldsymbol{Y}$ 是由 N 个均值为零、单位方差的不相关高斯随机变量组成的向量：$\mathcal{E}[\boldsymbol{Y}]=\boldsymbol{0}$，$\mathcal{E}[\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}]=\boldsymbol{I}$。设 $\boldsymbol{C}_{XX}$ 是希望的协方差阵(它也是自相关阵，因为均值为零)。由于 $\boldsymbol{C}_{XX}$ 是对称且正定的(PD)，它可用如下所示的楚列斯基分解法分解为(见附录 G)：

$$\boldsymbol{C}_{XX}=\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} \tag{6-223}$$

其中 $\boldsymbol{B}$ 是一个具有正对角元素的下三角矩阵。通过构造这个变换：

$$\boldsymbol{X}=\boldsymbol{B}\boldsymbol{Y} \tag{6-224}$$

我们得到需要的自相关矩阵：

$$\mathcal{E}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]=\boldsymbol{B}\,\mathcal{E}[\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}]\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{C}_{XX}=\boldsymbol{R}_{XX}\text{(对于零均值)} \tag{6-225}$$

图 6-28b 给出了一个相关高斯过程的现实示例，$R_{XX}(\tau)=\exp(-|\tau|/100)$。观察到这个现实比 6-28a 中的 iid 过程显得更容易“预期”。图 6-1 中的现实是这种相关高斯随机过程的另一个现实。

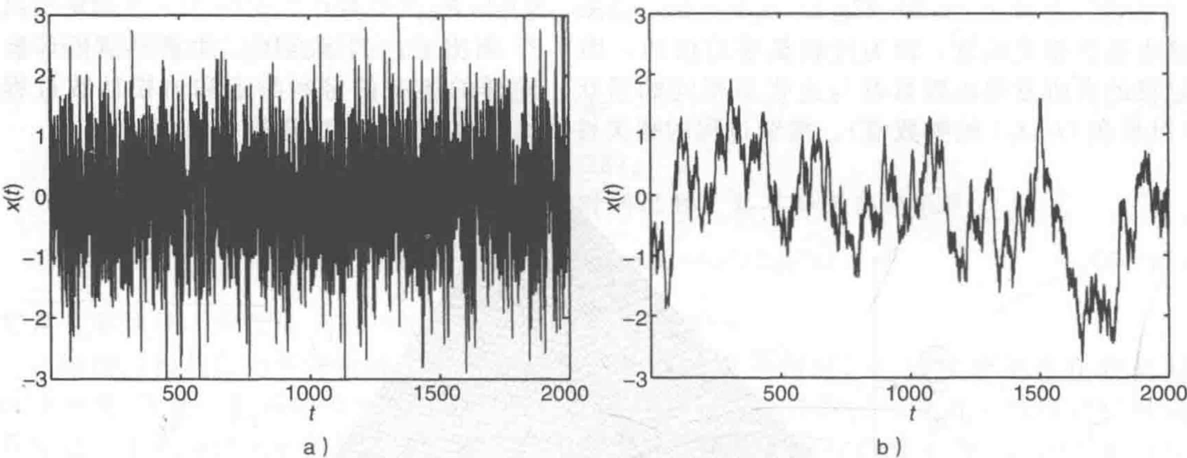


图 6-28　一个高斯过程的现实示例。a) 参数 $\{\mu=0,\sigma=1\}$ 的独立同分布(iid)；b) $\mu=0$，$\mathcal{E}[X(t_1)X(t_2)]=R_{XX}(t_2-t_1)=R_{XX}(\tau)-\exp(-|\tau|/100)$的广义平稳

6.12.5　脉冲幅度调制

随机过程 PAM 具有以下表述：

$$X(t)=\sum_{n=-\infty}^{\infty}A_n\operatorname{rect}((t-nT_0-T)/T_0-1/2) \tag{6-226}$$

其中 T_0 是一个幅度值的持续时间，$\{A_n\}$是 iid 随机变量，T 是$[0,T_0]$间均匀分布的随机变量，用来为所有实现的延迟时间建模。图 6-29 给出的是一个现实示例，其中 A_n 输出为$\{-3,-1,1,3\}$，$T_0=1$，$T=0$。因为是矩形函数，输出为±1 的 PAM 类似于随机电报信号，除了其具有更多结构：波形只能在符号周期 T_0 的整数倍时改变，且改变只能发生在 $A_n\neq A_{n+1}$时。这个随机信号的某些特性将在题 6-36 中进行分析。$\{A_n\}$的输出为±1 的特殊情况有时被称为随机二元信号。

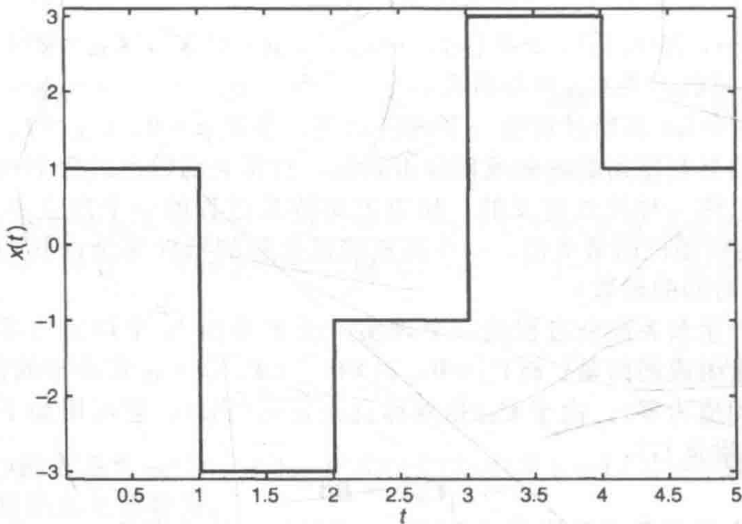


图 6-29　$T_0=1$，$T=0$ 的 4 阶 PAM 的实现示例 $x(t)$

6.12.6　随机正弦信号

最后，我们回到第 1 章中讨论的具有随机分量的正弦波形。这样的信号在通信系统中非常重要，通信系统中正弦波被随机过程调制后可在不同地方用高频载波来实现信息的传输。

考虑以下的随机过程：

$$X(t)=A\sin(2\pi f_c t+\Phi) \tag{6-227}$$

其中 A 为振幅，f_c 为载波频率，Φ 为相移，可用它们描述各种随机性。首先，假设 A 和 Φ 是不随时间改变的独立随机变量。例如，Φ 可以在 $[0,2\pi]$ 间均匀分布，A 可以是零均值、单位方差的高斯变量。当 A 和 Φ 的分布呈某一特殊值时可得到 $X(t)$ 的一个现实。图 6-30a给出了 $f_c=1\text{Hz}$ 时三个这样的现实，每个都有不同的振幅和相位。

随机正弦波可由允许振幅和相位为随机过程来产生：

$$X(t) = A(t)\sin(2\pi f_c t + \Phi(t)) \tag{6-228}$$

对于时变的情况，过程用一个 $A(t)$ 乘以正弦波，称做调幅(AM)。类似的，一个时变的相位 $\Phi(t)$ 产生调相(PM)信号(也可以变化频率 f_c，得到调频信号(FM))。图 6-30b和6-30c显示的是 AM 和 PM 的例子，其中 $A(t)$ 是振幅为 $\{1,2,3,4\}$ 的伯努利方案，$\Phi(t)$ 是相位为 $\{2\pi/5,4\pi/5,6\pi/5,8\pi/5\}$ 的伯努利方案，对应各自的四个符号。载频为 $f_c=10\text{Hz}$，振幅和相位可每隔 0.2s 变化一次(符号周期)。因为调制波形 $A(t)$ 和 $\Phi(t)$ 只有有限数目的值，它们是数字调制的例子，分别称为幅移键控(ASK)和相移键控(PSK)。注意到 6.11 节讨论过的 PAM 就是一个基带的 ASK，因为它不包括正弦载波。PAM 的情况类似，一个正弦波的数字调制具有更多结构，因为：(ⅰ)振幅和相位的数目是有限的；(ⅱ)它们只在一个确定的符号周期的整数倍处变化。这种结构在接收机中被用来确定传输的符号，第 10 章数字通信中将详细讨论这个问题。

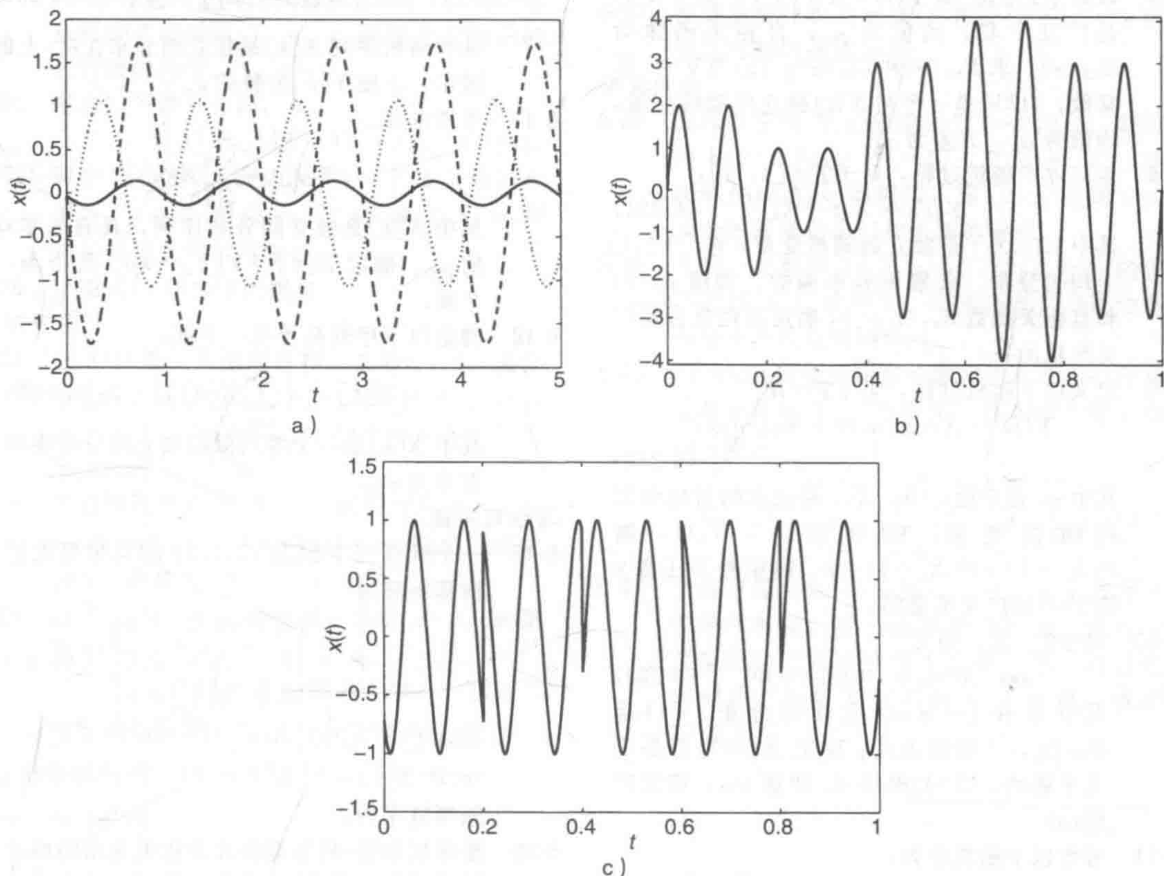


图 6-30 随机正弦波。a) 振幅和相位是随机变量的现实；b) 幅移键控(ASK)；c) 相移键控(PSK)

习题

基本随机过程

6-1 假设一个随机过程的样本空间为 $\mathcal{S}=\{1,2,3,4\}$，设对于 $\mathcal{T}=\mathcal{R}^+$ 有 $X(t,\zeta)=\exp(-\zeta t)u(t)$。对等可能的 ζ，详述在 $t=1$ 和 $t=2$ 时随机变量的概率空间 $\{\Omega,\mathcal{F},P\}$，画出所有现实。

6-2 设一个随机过程的样本空间为 $\mathcal{S}=[0,1]$，则对于 $\mathcal{T}=[0,5]$ 有 $X(t,\zeta)=\zeta(t-2)u(t-3)+\exp(-\zeta t)u(t)$。对于均匀分布的 ζ，求出 $t=$

2 和 $t=4$ 时随机变量的概率密度。

6-3　考虑以下随时间线性增长的随机过程：

$$Y(t)=at+X \tag{6-229}$$

其中 a 是一个常数，X 在 $[-1,1]$ 上均匀分布。求 $\mathcal{T}=\mathcal{R}^+$ 时 $Y(t)$ 的累积分布函数和概率密度。

6-4　设 $X(t)=A\exp(-t)u(t)$，其中 A 服从离散的均匀分布，输出为 $\{-1,0,1,2\}$。求 $\mathcal{T}\in\mathcal{R}^+$ 时 $X(t)$ 的概率密度。

6-5　考虑以下随机过程的积分：

$$Y(t)=\int_0^t X(\tau)\mathrm{d}\tau \tag{6-230}$$

设 $X(t)$ 与习题 6-4 中相同，除了 A 是一个标准高斯随机变量。求 $Y(t)$ 的概率密度和累积分布函数。

相关性和平稳性

6-6　求习题 6-3 中 $Y(t)$ 的均值、方差、自相关函数和自协方差函数。

6-7　设对于 $\mathcal{T}\in\mathcal{R}^+$ 有 $Y(t)=X(t)+V$，其中 $X(t)$ 是广义平稳，均值为 μ_X，自相关函数为 $R_{XX}(\tau)$。求 $R_{YY}(\tau)$ 和 $C_{YY}(\tau)$：(a)当 $V=v$ 是常数；(b)V 是一个与 $X(t)$ 独立的随机变量，均值为 μ_V，方差为 σ_V^2。

6-8　定义以下随机过程，对于 $\mathcal{T}=[0,1]$：

$$Y(t)=X_1t-X_2 \tag{6-231}$$

其中 $\{X_1,X_2\}$ 是独立的随机变量，在 $[-1,1]$ 上均匀分布。求累积分布函数、均值 $\mu_Y(t)$ 和自相关函数 $R_{YY}(t_1,t_2)$，确定过程是否是广义平稳的。

6-9　定义以下随机过程，对于 $\mathcal{T}=\mathcal{R}$：

$$Y(t)=X_1\cos(\omega_0t)+X_2\sin(\omega_0t) \tag{6-232}$$

其中 ω_0 是定值，$\{X_1,X_2\}$ 是独立的对称伯努利随机变量，输出为 $\{-1,1\}$，则 $P(X_1=1)=P(X_2=1)=p$。确定对于任意 p 值 $Y(t)$ 是广义平稳的。

6-10　对于 $\mathcal{T}=\mathcal{R}$，定义

$$X(t)=\cos(\omega_0t+\Phi) \tag{6-233}$$

其中 Φ 在 $[-\pi,\pi]$ 上均匀分布。(a)求 $R_{XX}(t_1,t_2)$ 的表达式，确定 $X(t)$ 是否是广义平稳的；(b)如果用 sin 代替 cos，重复问题(a)。

6-11　考虑以下随机序列：

$$Y_1[k]=X[k]+2X[k-1] \tag{6-234}$$

$$Y_2[k]=X[k]+(1/2)Y_2[k-1] \tag{6-235}$$

其中当 $T=\mathcal{Z}^+$ 时，$X[k]$ 是零均值 iid 序列。(a)确定 $Y_1[k]$ 和 $Y_2[k]$ 是否广义平稳；(b)求互相关函数 $\mathcal{E}[Y_1[k_1]Y_2[k_2]]$。

6-12　设 $X(t)$ 是广义平稳高斯过程，均值为零、自相关函数为 $R_{XX}(\tau)$。(a)证明对于所有 $t\in\mathcal{T}$，有 $f_{X(t)}(x)=f_X(x)$，也就是说它也是一阶平稳的；(b)确定 $X(t)$ 是否严平稳。

6-13　推导表 6-4 中广义伯努利随机游走的自相关和自协方差函数。

6-14　对定义在 $\mathcal{T}=\mathcal{N}$ 上的随机序列 $X[k]$，设轮流投掷硬币偶数次 k，投掷骰子奇数次 k。设掷硬币的输出为 $\{0,1\}$，掷骰子的输出为常规的 6 个数字。(a)推导 $R_{XX}[m_1,m_2]$ 的表达式；(b)求对所有 k，$P(X[k]=X[k-1])$ 的值。

6-15　考虑定义在 $\mathcal{T}=\mathcal{Z}^+$ 上的独立同分布标准高斯随机变量组成的随机序列 $X[k]$。设对于 $\mathcal{T}=\mathcal{N}$ 和 $\alpha\in\mathcal{R}$，$Y[k]=X[k]+\alpha X[k-1]$。(a)求 $R_{YY}[m]$ 的表达式；(b)确定 $Y[k]$ 是否为独立增量序列；(c)对于 $Z[k]=X[k]+\alpha X[k-2]$ 重复问题(a)和(b)。

鞅

6-16　考虑以下乘积序列：

$$Y[k]=\alpha\prod_{n=1}^{k}X[n] \tag{6-236}$$

其中随机序列 $X[k]$ 是独立同分布在 $\mathcal{R}^+$ 上的输出。求使 $Y[k]$ 为鞅的 α。

6-17　考虑求和

$$Y[k]=\sum_{n=1}^{k}X[n] \tag{6-237}$$

其中 $X[k]$ 是独立同分布序列，具有非零均值 μ_X。确定 $Z[k]=Y[k]-k\mu_X$ 是否为一个鞅。

6-18　确定以下序列是否是一个鞅：

$$Y[k]=\Big(\sum_{n=1}^{k}X[n]\Big)^2-k\sigma_X^2 \tag{6-238}$$

其中 $X[k]$ 是一个零均值的独立同分布序列，方差为 σ_X^2。

马尔可夫链

6-19　一个具有三个状态 $\{1,2,3\}$ 的马尔可夫链，转移矩阵为

$$\boldsymbol{P}=\begin{bmatrix}0.1 & 0.3 & 0.6\\ 0 & 0.5 & 0.5\\ 0.2 & 0.2 & 0.6\end{bmatrix} \tag{6-239}$$

求(a)$P(X[2]=3\mid X[0]=1)$；(b)$P(X[2]=3)$；(c)$P(X[1]=2\mid X[2]=3)$。假设初始状态是等概率的。

6-20　推导切尔曼-科尔莫洛夫方程更通用的形式

$$p_{mn}^{(r)}=\sum_{q=1}^{M}p_{mq}^{(s)}p_{qn}^{(r-s)} \tag{6-240}$$

对于 $1<s<r$。

6-21　一个马尔可夫链有三个状态 $\{1,2,3\}$，转移矩阵为：

$$\boldsymbol{P}=\begin{bmatrix}0.2 & 0.4 & 0.4\\ 0.1 & 0.2 & 0.7\\ 0.8 & 0.1 & 0.1\end{bmatrix} \tag{6-241}$$

对于初始状态 $\boldsymbol{p}[0]=[0.2,0.3,0.5]^{\mathrm{T}}$，求 $\boldsymbol{p}[k]$ 的表达式，确定稳态概率。

6-22 重复习题 6-21，此时马尔可夫链有四个状态 $\{1,2,3,4\}$，转移矩阵为

$$\boldsymbol{P}=\begin{bmatrix}0.1 & 0.9 & 0 & 0\\ 0.5 & 0.5 & 0 & 0\\ 0 & 0 & 0.4 & 0.6\\ 0 & 0 & 0.3 & 0.7\end{bmatrix} \tag{6-242}$$

设 $\boldsymbol{p}[0]=[0.1,\ 0.2,\ 0.5,\ 0.2]^{\mathrm{T}}$。

6-23 考虑以下具有输出 $\{-1,1\}$ 的马尔可夫链 $X[k]$ 的转移矩阵：

$$\boldsymbol{P}=\begin{bmatrix}0.1 & 0.9\\ 0.5 & 0.5\end{bmatrix} \tag{6-243}$$

求序列 $-1,\ -1,\ -1,\ -1,\ 1,\ -1$ 和 $1,\ 1,\ 1,\ 1,\ 1,\ -1$ 的概率。

6-24 假设我们希望扩展例 6-23 中伯努利比特计数器来对第一次出现 $\{1,0,1\}$ 进行检测。(a)推导这个方案的状态转移图；(b)用一步分析法求检测到 $\{1,\ 0,\ 1\}$ 序列之前收到的比特数的均值。

6-25 考虑以下差分方程：

$$Y[k]=\alpha Y[k-1]+X[k] \tag{6-244}$$

其中 $Y[0]=0$。设 $X[k]$ 为一个独立同分布序列，$k\geqslant 0$，假设 $|\alpha|<1$。确定 $X[k]$ 和 $Y[k]$ 是否为马尔可夫链。

6-26 证明式(6-131)中的概率。

随机过程

6-27 设 $X(t)$ 是一个泊松过程，参数 $\lambda=3$。求以下特性：(a) $P(X(2)>2)$；(b) $P(X(1)=1,\ X(2)=2)$；(c)直到 $X(t)=5$ 出现前的平均等待时间。

6-28 对泊松计数过程 $X(t)$，证明对于 $t_0<t$ 和 $n<N$，$P(X(t_0)=n\mid X(t)=N)$ 满足二项式分布，参数为 $\{N,p=t_0/t\}$.

6-29 设 $X(t)$ 为一个泊松过程，参数 $\lambda=2$。如果 $X(4)=3$，求：
(a) $t=2$ 后两次到达的概率；(b) $t=1$ 后一次到达的概率。

6-30 求随机电报信号 $Z(t)=WY(t)$ 的均值、方差、失真度和超峰度。

6-31 推导式(6-205)中随机电报信号 $Z(t)$ 的自相关函数。

6-32 推导式(6-206)中半随机电报信号 $X(t)$ 的相关系数。

6-33 证明维纳过程的概率密度满足式(6-216)中二阶 DE。

6-34 设 $X(t)$ 为标准维纳过程，方差为 $\alpha\varepsilon^2 t$，求过程 $Y(t)=|X(t)|$ 的分布。

6-35 对维纳过程，我们知道 $f_{X(t_1),X(t_2)}(x_1,\ x_2)=f_{X(t_1)}(x_1)f_{X(t_2)-X(t_1)}(x_2-x_1)$，因为它具有独立增量。写出 $X(1)$ 和 $X(2)$ 的联合概率密度的表达式。

6-36 设 $\{0,1\}$ 是式(6-226)中 PAM 过程的等可能振幅。假设 $T_0=0$，求 $\mu_X(t)$ 和 $R_{XX}(t_1,t_2)$ 表达式。为了求自相关函数，应该考虑 $\{t_1,\ t_2\}$ 可对应的期望是否在同一个符号内。

6-37 推导泊松计数过程 $X(t)$ 在条件 $X(t)=n$ 时到达时间的条件分布。

仿真作业

6-38 设随机序列 $X[k]$ 的输出为 $\{-1,0,1\}$，概率为 $P(0)=1/2$，$P(-1)=P(1)=1/4$。用 rand 生成 $X[k]$ 的现实，将 $[0,1]$ 区间分成三个区间来给出三种可能的输出。

6-39 生成一个参数 $\lambda=2$，$r=\{1/2,1,3/2\}$ 的伽马过程 $X(t)$ 的现实。可用 MATLAB 中伽马分布来确定间隔时间。

6-40 用 randn 和楚列斯基分解方法来产生一个相关高斯序列 $X[k]$ 的现实，其均值为零，自相关矩阵：

$$\boldsymbol{R}_{XX}=\begin{bmatrix}21/16 & 5/8 & 1/4\\ 5/8 & 21/16 & 5/8\\ 1/4 & 5/8 & 21/16\end{bmatrix} \tag{6-245}$$

6-41 仿真习题 6-21 中马尔可夫链。为了产生 $\boldsymbol{P}$ 的条件概率，用 rand 分段子区间。通过对足够长的时间内状态发生数计数，估计稳态概率，确定其与习题 6-21 中理论值的区别。

进一步阅读

以下参考资料提供更多关于随机过程各种特性的信息：Davenport(1970)，Doob(1953)，Larson 和 Shubert(1979)，Gray 和 Davisson(2004)，Hajek (2009)，Leon-Garcia (2008)，Papoulis (1965)，Parzen(1962)，Viniotis(1998)。

以下参考书提供一些特殊问题的深入探讨。马尔可夫过程：Hoel、Port 和 Stone(1972)，Karlin 和 Taylor (1975)，Taylor 和 Karlin (1984)．鞅：Athreya 和 Lahiri (2006)，Billingsley (1985)，Larson 和 Shubert(1979).

第7章 随机收敛、微积分和分解

7.1 引言

本章中，我们考虑随机过程的其他特性，这些特性深入阐述了随机过程的特征。包括随机序列 $X[k]$在 $k\to\infty$时的收敛性，以及随机过程 $X(t)$的对应特性：在 $t\in\mathcal{T}$处连续，其中$\mathcal{T}$是一个连续的时间集。还会给出随机过程的导数和积分的定义，以及微分方程(DE)。这些都非常有用，因为在第 8 章中，我们考虑信号和线性系统，它们都是用常系数的 DE 来建模的。有了这些知识储备，我们就可以检验线性系统是如何改变随机过程的概率特性的，特别是它的相关性。最后将描述随机信号是怎样分解为各分量之和，为原始过程提供更多信息。

考虑图 7-1 中泊松过程的现实。本章中，我们感兴趣的是过程连续性的不同类型，这些类型可能没有用于每个现实。即使图中显示的现实不是连续的，也有关于连续性的统计定义使得这个泊松过程被看做是连续的。各种连续的定义应用在一个单独的时刻 t 或对所有 $t\in\mathcal{T}$。类似的随机定义用在 $t\in\mathcal{T}$时刻的导数和某时间间隔$[t_0, t]\subset\mathcal{T}$的积分。后面即将证明，将这些随机定义用于随机过程的总体比用于各自的现实更有用。相同的方法也用于随机序列的收敛性，我们首先描述这个问题。

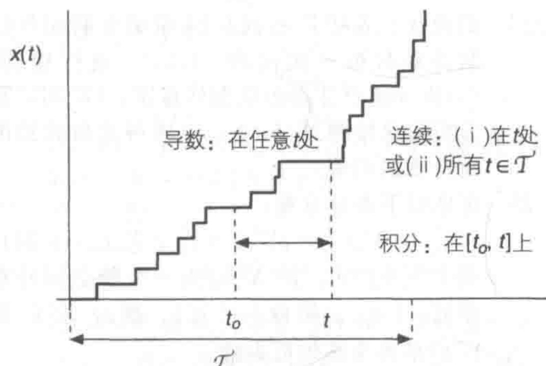


图 7-1　一个泊松随机过程的现实。过程的特性包括：在时刻 t 或整个 T 期间的连续性。t 时刻的导数，在时间间隔$[t_0, t]\subset\mathcal{T}$上的积分

7.2 随机收敛

随机序列收敛于一个常数或一个随机变量用于多个应用，比如参数估计(见第 9 章)。存在几种不同的收敛，它们的严格程度不同，但都提供随机序列在 $k\to\infty$时的特征信息。同时会引出大数定理，我们将在此章中讨论。

在第 2 章，我们描述了抽象概率空间 $\{\Omega,\mathcal{F},P\}$ 的几乎必然和必然事件。如果 $P(E)=1$ 或者如果 $P(E)=0$，事件 E 是几乎必然的。几乎必然并不意味着 E 会发生。如果事件 E 一定发生并且没有其他事件会发生，则 E 是必然的。同样，$P(E)=0$ 的零概率事件 E(但可能发生)和一个不可能事件(绝不会发生的事件)是有区别的。

例 7-1 考虑一个输出为 $\{0,1\}$ 的伯努利随机序列 $X[k](k\in\mathcal{Z}^+)$，$P(X[k]=0)=P(X[k]=1)=0.5$。一个无穷序列为全 1 的概率显然是零，尽管这样一个序列有可能出现。对某些 k，当 $k\to\infty$时 $X[k]=0$ 这个事件是几乎必然的；它也可能不发生，因此它不是必然的。◀

对例 7-1，我们看到之所以会引入几乎必然事件的概念是在进行了无限次观测后。我们将此定义用于随机序列 $X[k]$的收敛性(通常也用在随机变量序列 $\{X_1, X_2,\cdots\}$ 中)。

定义(几乎必然收敛)　随机序列 $X[k]$**几乎必然收敛**，如果

$$P(\lim_{k\to 0} X[k]=X)=1 \tag{7-1}$$

其中 X 是一个随机变量或一个常数。这种收敛也被称为**依概率 1 收敛**。常用以下的标记来

表示：$X[k]\xrightarrow{a.s}X$。用随机变量序列的基本样本空间$\mathcal{S}$(见第 6 章)，我们可将其表示的更精确：

$$P(\zeta\in\mathcal{S}:\lim_{k\to 0}X[k,\zeta]=X(\zeta))=1 \tag{7-2}$$

其中 $X(\zeta)$是一个随机变量或常数。

如果一个现实不收敛，则要么它发散为$\pm\infty$，要么振荡，且不会停止。

定义(收敛的现实)　现实 $x[k,\zeta]$ 收敛于 $x(\zeta)$，如果对每个 $\varepsilon>0$，存在一个 k 值，服从

$$|x[k,\zeta]-x(\zeta)|<\varepsilon \tag{7-3}$$

如果对所有 $\zeta\in\mathcal{S}$，$x(\zeta)$都是相同的值，则序列 $X[k]$收敛于一个常数。图 7-2a 对三个现实的情况进行了说明，它们都收敛于同一个数(纵轴上一个间隔为 2ε 区间的中心处)。它是一个分布由$\zeta\in\mathcal{S}$决定的随机变量：特定的ζ会给出特定的$x(\zeta)$的值。见图 7-2b，图中两个现实收敛于一个值，第三个现实收敛于随机变量的一个不同输出。

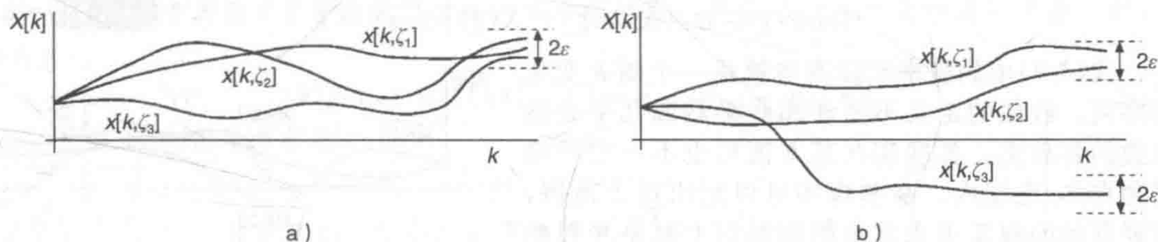


图 7-2　收敛现实的说明。a) 收敛到一个常数；b) 收敛到一个随机变量。(为了看起来更清晰采用了序列的连续模式)

定义(必然收敛)　随机序列 $X[k]$**必然收敛**，如果对所有 $\zeta\in\mathcal{S}$，有

$$\lim_{k\to\infty}X[k,\zeta]=X(\zeta) \tag{7-4}$$

其中 $X(\zeta)$是一个随机变量或一个常数。

对必然收敛，定义中并不包括概率，随机序列的所有现实都收敛，即使是零概率事件。必然收敛显然要强于几乎必然收敛，后者“忽略”了零概率非收敛的那些现实。但是，我们通常较少考虑必然收敛，因为在实际中我们一般不关注具有零概率的现实。

例 7-2　设一个随机序列的现实描述如下：

$$X[k,\zeta]=\zeta^k u[k] \tag{7-5}$$

其中 $u[k]$是离散单位阶跃函数，样本空间$\mathcal{S}$的输出是在 $[0,1]$ 上均匀分布的。(回顾上一章：$\mathcal{S}$的元素实际上是各个现实。但是，对某些具有如本例中所述结构的随机过程，通常有可能用一个随机变量或随机向量的输出来描述那些现实。)如果 $\zeta=0.5$，则对应的现实为 $X[k,0.5]=(0.5)^k u[k]$，它收敛于 0。显然 $X[k,\zeta]$ 对所有 $\zeta\in(0,1)$收敛于 0；只有 $X[k,1]$“收敛”于 1(对所有 k，它实际上是一个常数)。随机变量 $X[\zeta]$在收敛处是输出为 $\{0,1\}$ 的伯努利，其中 $P(X(\zeta)=1)=p=0$，$P(X(\zeta)=0)=q\triangleq 1-p=1$。因此，$X[\zeta]$是一个几乎必然是常数(等于 0)的随机变量的例子。由于所有的收敛值都是有限的，$X[k]$必然收敛。假设代之以：

$$X[k,\zeta]=k\zeta^k u[k] \tag{7-6}$$

这个随机序列不是必然收敛的，因为它在 $\zeta=1$ 时发散。但是，它是几乎必然收敛的，因为 $P(\zeta=1)=0$，对 $\zeta\in(0,1)$，在 $k\zeta^k$ 的乘积中 $\zeta^k\to 0$ 要快于 $k\to\infty$。◀

例 7-3　设一个随机序列的现实为：

$$X[k,\zeta]=\exp(\zeta k)u[k] \tag{7-7}$$

其中三个输出 $\mathcal{S}=\{-1,0,1\}$ 等概率。显然，$X[k]$不是几乎必然收敛的，因为当$\zeta=1$(具有非零概率)时它发散。同样，如果$\mathcal{S}$的输出是在 $[-1,1]$ 上均匀分布的，$X[k]$不是几乎必

然收敛的。与 $\zeta<1$ 对应的现实以概率 $P(\lim_{k\to\infty} X[k]=0)=1/2$ 收敛于 0。与 $\zeta=0$ 对应的现实"收敛"于 1，所有其他现实(联合概率为 1/2)是发散的。◀

几乎必然收敛意味着随机序列的每个具有非零概率的现实收敛到一个常数。这个常数对所有现实不一定都相同，某种情况序列仍收敛，但收敛到一个随机变量。由于几乎必然收敛应用于每个现实，相比于其他类型的收敛而言，它是一个很强的条件。例如，依概率收敛是一种较弱的收敛类型。

定义(依概率收敛)　随机序列 $X[k]$ **依概率收敛**，如果对每个 $\varepsilon>0$ 有

$$\lim_{k\to\infty} P(|X[k]-X|>\varepsilon)=0 \tag{7-8}$$

其中 X 是一个随机变量或一个常数。通常用以下符号表示：$X[k]\xrightarrow{p}X$。

与式(7-1)比较，我们看到这里的取极限是在求概率表达式之外。与式(7-2)对应的依概率收敛为：

$$\lim_{k\to\infty} P(\zeta\in\mathcal{S}:|X[k,\zeta]-X(\zeta)|>\varepsilon)=0 \tag{7-9}$$

式(7-8)中的概率实际本身就是一个随 k 变化的序列。收敛的定义不关注那些必然或几乎必然收敛的各现实。各现实在某 k 值后也不一定严格限制在 2ε 范围内。某些现实可以超出这个范围，只要其他的现实进入此范围满足以上概率序列的极限值，如图 7-3 所示。

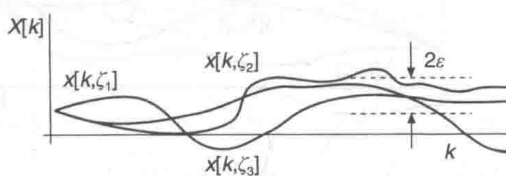


图 7-3　依概率收敛的图解(为了看起来更清晰采用了序列的连续现实)

例 7-4　考虑以下随机序列：

$$Y[k]=\frac{1}{k}\sum_{m=1}^{k}X[m] \tag{7-10}$$

其中 $\{X[m]\}$ 是 iid 高斯随机变量，参数为 $\{\mu=0,\sigma\}$。显然 $Y[k]$ 也是零均值高斯的，但具有递减方差 σ^2/k。这个方差很容易计算，因为有独立性的假设(见第 5 章)。在接下来的推导中，我们将证明 $Y[k]$ 依概率收敛于 0。由定义，

$$P(|Y[k]-0|>\varepsilon)=P(Y[k]>\varepsilon)+P(Y[k]\leqslant-\varepsilon)=2P(Y[k]>\varepsilon) \tag{7-11}$$

因为 $Y(k)$ 的 pdf 是偶数，上式被简化。由第 3 章可知

$$\begin{aligned}P(Y[k]>\varepsilon)&=\frac{1}{\sqrt{2\pi(\sigma^2/k)}}\int_{\varepsilon}^{\infty}\exp(-v^2/2(\sigma^2/k))\mathrm{d}v\\&=\frac{1}{\sqrt{2\pi}}\int_{\varepsilon/(\sigma/\sqrt{k})}^{\infty}\exp(-v^2/2)\mathrm{d}v=Q(\sqrt{k}\varepsilon/\sigma)\end{aligned} \tag{7-12}$$

它为 Q 函数(见附录 B)。正如上面提到的，这个概率是一个确定的序列而且不关注各单独的现实。对于依概率收敛，我们只需要式(7-12)对每个 $\varepsilon>0$ 收敛到 0 即可。由 Q 函数的特性，即：

$$\lim_{k\to\infty} Q(\sqrt{k}\varepsilon/\sigma)=0 \tag{7-13}$$

因此 $Y[k]$ 依概率收敛到一个常数。◀

另一种不关注各单独的现实的收敛是依均值收敛。

定义(依均值收敛)　随机序列 $X[k]$ **依均值收敛**，如果满足

$$\lim_{k\to\infty} \mathcal{E}[|X[k]-X|^m]=0 \tag{7-14}$$

其中 X 是一个随机变量或一个常数，$m\in\mathcal{N}$。通常用以下符号表示：$X[k]\xrightarrow{L^m}X$，其中 L^m 是 L^m 范数(当 $m=2$ 时是欧几里得范数)。

可以证明：L^m 收敛意味着对 $m>n$，L^n 也收敛(见题 7-4)。L^m 收敛应用最广泛的形式是 $m=2$ 时对应的均方(MS)收敛；以下符号也用于这种情况：$X[k]\xrightarrow{\text{m. s.}}X$。以上的期望值

是一个趋于确定的序列，是式(7-8)中的概率。它不关注各单独的现实，只关注总体的 $|X[k]-X|^m$ 趋于 0。显然这种收敛要强于依概率收敛，稍后将加以证明。

例 7-5 对例 7-4 中的随机序列，我们需要证明对 MS 收敛有：

$$\lim_{k\to\infty}\varepsilon[Y^2[k]]=0 \tag{7-15}$$

这很容易得到，因为 $Y[k]$ 具有零均值，而且我们已经知道 $Y[k]$ 的方差为 σ^2/k。

考虑随机序列 $X[k]$，对某个特定的时刻 k，cdf 为 $F_{X[k]}(x)=P(X[k]\leqslant x)$。我们再定义一种收敛，它比之前所有收敛的定义都要弱。

定义(依分布收敛) 随机序列 $X[k]$ **依分布收敛**，如果

$$\lim_{k\to\infty}F_{X[k]}(x)=F_X(x) \tag{7-16}$$

对所有 x 假设 cdf$F_X(x)$ 存在。常采用以下符号标记：$X[k]\xrightarrow{d}X$。◀

例 7-6 考虑例 6-2 中的随机过程，其分布如下：假设 T 为一个离散随机变量，则 $t\to k\in\mathcal{Z}^+$：

$$F_{Y[k]}(y)=u(y)F_T[k]+F_X(y)(1-F_T[k]) \tag{7-17}$$

显然 $Y[k]$ 依分布收敛：

$$\lim_{k\to\infty}F_{Y[k]}(y)=u(y) \tag{7-18}$$

因为 $F_T[k]\to1$，对应的 pdf 为 $\delta(y)$，意味着 $Y[k]$ 实际依分布收敛于 0。◀

表 7-1 总结了 5 种收敛类型，图 7-4 解释了它们的相互关系。

表 7-1 随机序列 $X[k]$ 收敛的总结

收敛类型	定义
必然：$X[k]\xrightarrow{s}X$	$\lim_{k\to\infty}X[k,\zeta]=X(\zeta)$，对所有的 $\zeta\in S$
几乎必然：$X[k]\xrightarrow{a.s}X$	$P(\lim_{k\to\infty}X[k]=X)=1$
依概率：$X[k]\xrightarrow{p}X$	$\lim_{k\to\infty}P(\lvert X[k]-X\rvert>\varepsilon)=0$，对每一个 $\varepsilon>0$
依均值：$X[k]\xrightarrow{L^m}X$	$\lim_{k\to\infty}\mathcal{E}[\lvert X[k]-X\rvert^m]=0$，$m\in\mathcal{N}$
依分布：$X[k]\xrightarrow{d}X$	$\lim_{k\to\infty}F_{X[k]}(x)=F_X(x)$

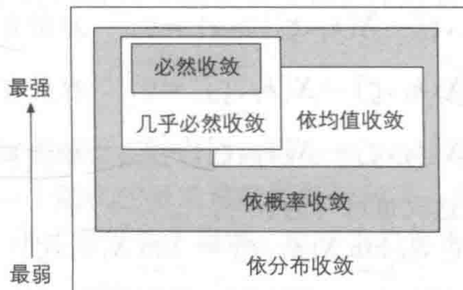


图 7-4 随机序列的五种类型的收敛之间的相互关系：$s\subseteq a.s.\subseteq p\subseteq d$，$L^m\subseteq p\subseteq d$。通常，我们不能对必然收敛、几乎必然收敛与依均值收敛的相互关系下结论

定理 7-1 依均值收敛意味着依概率收敛。

证明： 由附录 F 中的比安内梅(Bienaymé)不等式，用 $X[k]-X$ 代替 X，$\alpha=0$，可以得到：

$$P(|X[k]-X|\geqslant\varepsilon)\leqslant(1/\varepsilon^2)\,\mathcal{E}[|X[k]-X|^m] \tag{7-19}$$

对每个 $\varepsilon>0$，设 $k\to\infty$，则

$$\lim_{k\to\infty}P(|X[k]-X|\geqslant\varepsilon)\leqslant(1/\varepsilon^2)\lim_{k\to\infty}\mathcal{E}(|X[k]-X|^m] \tag{7-20}$$

假设依均值收敛，则不等号右边为 0，证明完毕。

定理 7-2 MS 收敛意味着均值和方差也收敛：

$$X[k] \xrightarrow{\text{m. s.}} X \Rightarrow \lim_{k\to\infty} \mathcal{E}[X[k]] = \mathcal{E}[X] \quad \lim_{k\to\infty} \operatorname{var}[X[k]] = \operatorname{var}[X] \tag{7-21}$$

证明： 均值：由于期望运算是线性的，$|\mathcal{E}[X[k]] - \mathcal{E}[X]| = |\mathcal{E}[X[k] - X]|$，由柯西-施瓦茨不等式(见附录 F，$g(X) = X[k] - X$ 和 $h(X) = 1$)：

$$|\mathcal{E}[X[k] - X]| \leqslant \sqrt{\mathcal{E}[(X[k] - X)^2]} \tag{7-22}$$

如果最后一个表达式趋于 0(即 MS 收敛)，则$\mathcal{E}[X[k]]$收敛于$\mathcal{E}[X]$。

方差：由三角不等式(见附录 F)：

$$\sqrt{\mathcal{E}[X^2[k]]} = \sqrt{\mathcal{E}[(X[k] - X + X)^2]} \leqslant \sqrt{\mathcal{E}[(X[k] - X)^2]} + \sqrt{\mathcal{E}[X^2]} \tag{7-23}$$

将最后一个式中的最后一项移到不等号的左边得到：

$$\sqrt{\mathcal{E}[X^2[k]]} - \sqrt{\mathcal{E}[X^2]} \leqslant \sqrt{\mathcal{E}[(X[k] - X)^2]} \tag{7-24}$$

如果右边趋于 0，则二阶矩$\mathcal{E}[X^2[k]]$收敛于$\mathcal{E}[X^2]$，$\mathcal{E}[X^2[k]] - \mathcal{E}^2[X[k]] = \operatorname{var}[X[k]]$收敛于 $\operatorname{var}[X]$。

由于随机变量 X 是否收敛是未知的，以下定理可用来证明 MS 收敛。

定理 7-3(柯西准则) 随机序列 $X[k]$依 MS 收敛，当且仅当：

$$\lim_{k_1, k_2\to\infty} \mathcal{E}[(X[k_1] - X[k_2])^2] = 0 \tag{7-25}$$

证明： 我们只证明定理的必要条件，将式(7-25)重写如下：

$$\begin{aligned}\mathcal{E}[(X[k_1] - X + X - X[k_2])^2] =& \mathcal{E}[(X[k_1] - X)^2] + 2\,\mathcal{E}[(X[k_1] - X)(X - X[k_2])] \\ &+ \mathcal{E}[(X - X[k_2])^2]\end{aligned} \tag{7-26}$$

由柯西-施瓦茨不等式，右边式子的中间项受以下条件约束：

$$\begin{aligned}\mathcal{E}[(X[k_1] - X)(X - X[k_2])] &\leqslant |\mathcal{E}[(X[k_1] - X)]\,\mathcal{E}[(X - X[k_2])]| \\ &\leqslant \sqrt{\mathcal{E}[(X[k_1] - X)^2]\,\mathcal{E}[(X - X[k_2])^2]}\end{aligned} \tag{7-27}$$

则，如果$\mathcal{E}[(X[k_1] - X)^2] \to 0$，$\mathcal{E}[(X - X[k_2])^2] \to 0$，式(7-25)中的极限成立。

在以上 MS 意义下用到的柯西准则是实际分析中对确定性情况的扩展，其中一个数值序列$\{x_k\}$收敛，当且仅当对子序列$\{x_{k_1}\}$和$\{x_{k_2}\}$满足：

$$\lim_{k_1, k_2\to\infty} |x_{k_1} - x_{k_2}| = 0 \tag{7-28}$$

柯西准则也可用于其他形式的收敛，我们总结如下：

依概率：$\lim\limits_{k_1, k_2\to\infty} P(|X[k_1, \zeta] - X[k_2, \zeta]| > \varepsilon) = 0$，对所有的 $\varepsilon > 0$ (7-29)

必然：$\lim\limits_{k_1, k_2\to\infty} |X(k_1, \zeta] - X[k_2, \zeta]| = 0$，对所有的 $\zeta \in \mathcal{S}$ (7-30)

几乎必然：$\lim\limits_{k_1, k_2\to\infty} |X[k_1, \zeta] - X[k_2, \zeta]| = 0$，对所有的 $\zeta \in E \subset \mathcal{S}$ (7-31)

其中 $P(E) = 1$。最后一个表达式也可写为：

$$P(\zeta \in \mathcal{S}: \lim_{k_1\to\infty} X[k_1, \zeta] = \lim_{k_2\to\infty} X[k_2, \zeta]) = 1 \tag{7-32}$$

以下的定理对鞅序列有用，我们不加证明地给出。

定理 7-4(鞅收敛) 如果 $X[k]$是一个鞅序列，满足$\mathcal{E}[X^2[k]] < \infty$，则当 $k \to \infty$ 时，它几乎必然收敛于随机变量 X 或一个常数。

接下来，我们给出 MS 收敛的两个额外的特性来强调定理 7-2 和定理 7-3 中用到的理论。在证明过程中我们的目标是重写感兴趣的序列的期望值使其受限于已知的依 MS 收敛的其他序列的期望值。通常可通过柯西-施瓦茨不等式或三角不等式得到边界，正如上面定理中用到的。接下来给出证明(Stark 和 Woods，2002)。

- 用一个随机变量相乘 设 $X[k]$依 MS 收敛到 X，Y 为一个随机变量，假设它们有有限的二阶矩：$\mathcal{E}[X^2[k]] < \infty$和$\mathcal{E}[Y^2] < \infty$，则

$$\lim_{k\to\infty}\mathcal{E}[X[k]Y]=\mathcal{E}[XY] \tag{7-33}$$

减去 XY，用柯西不等式得到：

$$|\mathcal{E}[X[k]Y-XY]|\leqslant|\mathcal{E}[X[k]-X]||\mathcal{E}[Y]|\leqslant\sqrt{\mathcal{E}[(X[k]-X)^2]\mathcal{E}[Y^2]} \tag{7-34}$$

由于 $X[k]\xrightarrow{\text{m.s.}}X$ 并且$\mathcal{E}[Y^2]$是有限的，不等式的右边趋于 0，$X[k]Y\xrightarrow{\text{m.s.}}XY$。

- *用随机序列相乘*　设 $X[k]$如上所示，假设 $Y[k]$依 MS 收敛于 Y。我们通常不能说 $X[k]Y[k]$依 MS 收敛于 XY。由三角不等式：

$$\begin{aligned}\sqrt{\mathcal{E}[(X[k]Y[k]-XY)^2]}&=\sqrt{\mathcal{E}[(X[k](Y[k]-Y)+(X[k]-X)Y)^2]}\\&\leqslant\sqrt{\mathcal{E}[X^2[k](Y[k]-Y)^2]}+\sqrt{\mathcal{E}[(X[k]-X)^2Y^2]}\end{aligned} \tag{7-35}$$

其中在第一行中加上和减去了 $X[k]Y$ 项。由于期望中的乘积使每个序列各自的 MS 收敛没法用。注意如果 $X[k]$和 $Y[k]$是不相关的，则期望值可分解为：

$$\sqrt{\mathcal{E}[(X[k]Y[k]-XY)^2]}\leqslant\sqrt{\mathcal{E}[X^2[k]]}\sqrt{\mathcal{E}[[Y[k]-Y)^2]}+\sqrt{\mathcal{E}[(X[k]-X)^2]}\sqrt{\mathcal{E}[Y^2]} \tag{7-36}$$

这是一个不太重要的结论。对于一般的情况，我们由柯西-施瓦茨不等式得到：

$$\mathcal{E}[X^2[k](Y[k]-Y)^2]\leqslant\sqrt{\mathcal{E}[X^4[k]]}\sqrt{\mathcal{E}[[Y[k]-Y)^4]} \tag{7-37}$$

$$\mathcal{E}[(X[k]-X)^2Y^2]\leqslant\sqrt{\mathcal{E}[(X[k]-X)^4]}\sqrt{\mathcal{E}[Y^4]} \tag{7-38}$$

为了用到 $X[k]$和 $Y[k]$的 MS 假设，我们需要不等式右边的二阶矩(不是四阶矩)。但是，如果 $X[k]$和 $Y[k]$是高斯的，则由第 5 章可知四阶矩可由二阶矩得到：

$$\mathcal{E}[X^4[k]]=3\,\mathcal{E}^2[X^2[k]],\quad \mathcal{E}[(X[k]-X)^4]=3\,\mathcal{E}^2[(X[k]-X)^2] \tag{7-39}$$

对 $Y[k]$和 Y 有同样的结论。由于 $X[k]\xrightarrow{\text{m.s.}}X$，$Y[k]\xrightarrow{\text{m.s.}}Y$，它们的二阶矩是有限的，对高斯序列式(7-35)的右边极限值趋于零。如果 $X[k]\xrightarrow{L^4}X$，$Y[k]\xrightarrow{L^4}Y$(这是一个强于 MS 收敛的条件)，则 $X[k]Y[k]\xrightarrow{L^2}XY$ 不需要高斯假设。

以下的附加特性在例 7-6 中说明。

- *随机序列的和*　如果 $X[k]\xrightarrow{\text{m.s.}}X$，$Y[k]\xrightarrow{\text{m.s.}}Y$，则 $X[k]+Y[k]\xrightarrow{\text{m.s.}}X+Y$
- *互相关收敛*　对 $X[k]$和 $Y[k]$ MS 收敛，如上行的条件所示(Hajek，2009)：

$$\lim_{k\to\infty}\mathcal{E}[X[k]Y[k]]=\mathcal{E}[XY] \tag{7-40}$$

这个结果与定理 7-2 中的均值和方差收敛类似。注意式(7-40)与式(7-35)中 $X[k]Y[k]$的 MS 收敛不同。

7.3　大数定理

大数定理描述了随机序列的特定函数在时间趋向于无穷大时的表现。它们是之前收敛定义中关于样本均值收敛到一个随机变量或常数的特殊情况。考虑 iid 随机序列 $X[k]$，均值为 μ_X，有限方差 $\sigma_X^2<\infty$。定义以下样本均值，此样本均值本身就是一个随机序列：

$$\overline{X}[k]\triangleq\frac{1}{k}\sum_{m=1}^{k}X[m] \tag{7-41}$$

并且是总体均值 μ_X 的一个估计(估计器及其特性在第 9 章中讨论)。

定理 7-5(*弱大数定理*)　*样本均值估计器* $\overline{X}[k]$*依概率收敛于* μ_X：

$$\lim_{k\to\infty}P(|\overline{X}[k]-\mu_X|\geqslant\varepsilon)=0 \tag{7-42}$$

对每个 $\varepsilon>0$。

证明：由于假设 $X[k]$是一个 iid 序列，样本均值的期望值为：

$$\mathcal{E}[\overline{X}[k]]=\frac{1}{k}\sum_{m=1}^{k}\mathcal{E}[X[m]]=\frac{1}{k}k\mu_X=\mu_X \tag{7-43}$$

因此它是一个无偏估计器，它的方差为：

$$\mathcal{E}[(\overline{X}[k]-\mu_X)^2]=\mathcal{E}\left[\left(\frac{1}{k}\sum_{m=1}^{k}(X[m]-\mu_X)\right)^2\right]=\frac{1}{k^2}\sum_{m=1}^{k}\sigma_X^2=\sigma_X^2/k \tag{7-44}$$

由附录 F 中的切比雪夫不等式得到：

$$P(|\overline{X}[k]-\mu_X|\geqslant\varepsilon)\leqslant\frac{\sigma_X^2}{k\varepsilon^2} \tag{7-45}$$

其中代入了上式的方差。这样，对每个 $\varepsilon>0$

$$\lim_{k\to\infty}P(|\overline{X}[k]-\mu_X|\geqslant\varepsilon)\leqslant\lim_{k\to\infty}\frac{\sigma_X^2}{k\varepsilon^2}=0 \tag{7-46}$$

证明完毕。

正如之前讨论的，这个概率是一个确定的序列。它是一个“弱定理”，因为它是基于依概率收敛。强大数定理指的是样本均值的几乎必然收敛，这是一个更强的条件，因为它关注到了各单独的现实。要证明强大数定理，克罗内克引理是一个有用的结论，我们将不加证明的给出。

引理 7-1（克罗内克） 假设非随机序列 $x[k](k\in\mathcal{N})$ 收敛于 c，如下式所示：

$$\sum_{k=1}^{\infty}x[k]/a_k=c<\infty \tag{7-47}$$

其中 $0=a_0<a_1\leqslant a_2\leqslant\cdots$ 是一个递增的系数集，则

$$\lim_{k\to\infty}(1/a_k)\sum_{m=1}^{k}x[m]=0 \tag{7-48}$$

定理 7-6（强大数定理） 只要满足

$$P(\lim_{k\to\infty}\overline{X}[k]=\mu_X)=1 \tag{7-49}$$

样本均值估计器 $\overline{X}[k]$ 几乎必然收敛于 μ_X。

证明：(Larson 和 Shubert，1979)定义零均值序列

$$Y[k]\triangleq\sum_{m=1}^{k}(X[m]-\mu_X)/m \tag{7-50}$$

它的方差为：

$$\mathrm{var}[Y[k]]=\sum_{m=1}^{k}(1/m^2)\,\mathcal{E}[(X[m]-\mu_X)^2]=\sigma_X^2\sum_{m=1}^{k}(1/m^2) \tag{7-51}$$

其中我们用到了 $X[k]$ 是一个 iid 序列的事实。这个方差是有限的，因为求和项是收敛的(见附录 E)。观察到 $Y[k]$ 是一个鞅

$$\begin{aligned}\mathcal{E}[Y[k]|Y[1],\cdots,Y[k-1]]&=\mathcal{E}\Big[\sum_{m=1}^{k}(1/m)(X[m]-\mu_X)|Y[1],\cdots,Y[k-1]\Big]\\&=\mathcal{E}[(1/k)(X[k]-\mu_X)+Y[k-1]|Y[1],\cdots,Y[k-1]]\\&=(1/k)\,\mathcal{E}[X[k]-\mu_X]+Y[k-1]=Y[k-1]\end{aligned} \tag{7-52}$$

利用之前的鞅收敛定理(见定理 7-4)，当 $k\to\infty$ 时，$Y[k]$ 几乎必然收敛于随机变量 Y(或一个常数)。由于 $Y[k]$ 的一个现实具有以下形式：

$$y[k]=\sum_{m=1}^{k}(x[m]-\mu_X)/m \tag{7-53}$$

我们可以采用克罗内克引理，得到：

$$\lim_{k\to\infty}\frac{1}{k}\sum_{m=1}^{k}(x[m]-\mu_X)=0 \tag{7-54}$$

引理中 $a_k=k$。则几乎所有具有 $\bar{x}[k]=(1/k)\sum_{m=1}^{k}x[m]$ 这种形式的现实收敛于 μ_X，证明完毕。

图 7-5 显示的是 iid 标准高斯序列 $X[k]$ 的样本均值的现实的例子，证明了它们趋近于零。

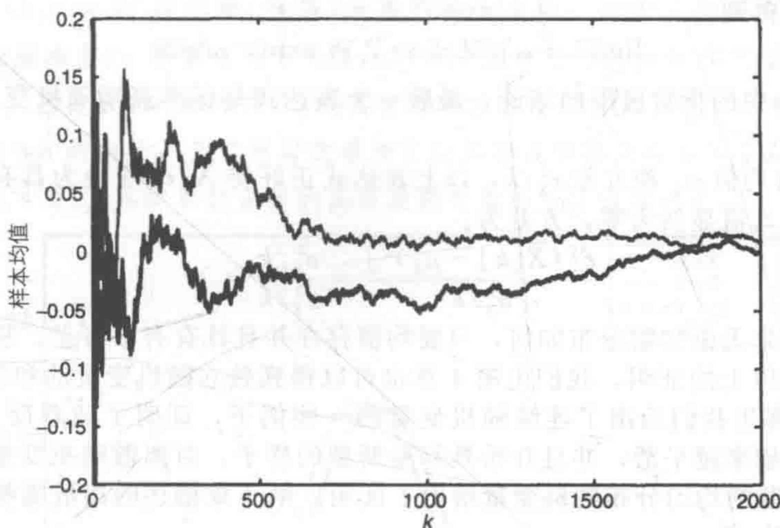


图 7-5　一个 iid 标准高斯随机序列的样本均值 $\overline{X}[k]$ 的现实的例子

7.4 中心极限定理

中心极限定理(CLT)描述了独立随机变量和的一个重要特性。

定理 7-7(中心极限)　设 $X[k]$ 是一个 iid 随机序列，均值为 μ_X，方差为 σ_X^2，定义以下样本均值的函数：

$$Y[k] \triangleq \frac{\overline{X}[k]-\mu_X}{\sigma_X/\sqrt{k}} \tag{7-55}$$

则 $\lim\limits_{k\to\infty} Y[k]$ 具有标准高斯分布。

这个定理指的是特定的依分布收敛。

证明：样本均值的特征函数为：

$$\Phi_{\overline{X}[k]}(\omega)=\mathcal{E}[\exp(\mathrm{j}\omega\overline{X}[k])]=\Phi_{\sum\limits_{n=1}^{k}X[n]}(\omega/k)=\prod_{n=1}^{k}\Phi_X(\omega/k)=\Phi_X^k(\omega/k) \tag{7-56}$$

其中 $\sum\limits_{n=1}^{k}X[n]$ 的比值因子 $1/k$ 和 ω 组合在一起，由 iid 假设得到最后的结果。可以直接证明对式(7-55)的变换(见例 7-13)：

$$\Phi_{Y[k]}(\omega)=[\exp(-\mathrm{j}\omega\mu_X/\sqrt{k}\sigma_X)\Phi_X(\omega/\sqrt{k}\sigma_X)]^k \tag{7-57}$$

为了标记方便，定义 $\omega_k \triangleq \omega/\sqrt{k}\sigma_X$，得到以下对指数的级数展开：

$$\Phi_X(\omega_k)=\mathcal{E}[\exp(\mathrm{j}\omega_k X)]=\mathcal{E}\Big[\sum_{n=0}^{\infty}\frac{(\mathrm{j}\omega_k X)^n}{n!}\Big]=\sum_{n=0}^{\infty}\frac{(\mathrm{j}\omega_k)^n\,\mathcal{E}[X^n]}{n!}$$

$$=1+\mathrm{j}\omega_k\,\mathcal{E}[X]-\omega_k^2\,\mathcal{E}[X^2]/2+\sum_{n=3}^{\infty}\frac{(\mathrm{j}\omega_k)^n\,\mathcal{E}[X^n]}{n!} \tag{7-58}$$

其中求和项中的前三项被分离出来。同样对式(7-57)中的指数有：

$$\exp(-\mathrm{j}\omega_k\mu_X)=\sum_{n=0}^{\infty}\frac{(-\mathrm{j}\omega_k\mu_X)^n}{n!}=1-\mathrm{j}\omega_k\mu_X-\omega_k^2\mu_X^2/2+\sum_{n=3}^{\infty}\frac{(-\mathrm{j}\omega_k\mu_X)^n}{n!} \tag{7-59}$$

将 $\mathcal{E}[X^2]=\sigma_X^2+\mu_X^2$ 代入式(7-58)，将以上两个级数相乘，发现 $n\in\{0,1,2\}$ 的大多数乘积项消掉了，因此：

$$\Phi_{Y[k]}(\omega)=[1-\omega_k^2\sigma_X^2/2+\cdots]^k \tag{7-60}$$

其中…标出的是指数 ω_k^n 超过 $n=2$ 的高阶项。当 $k\to\infty$ 取极限时，由于 ω_k 分母中的 $\sqrt{k}$，这

些项可以忽略，得到：

$$\lim_{k\to\infty}[1-\omega^2/2k+\cdots]^k=\exp(-\omega^2/2) \tag{7-61}$$

其中用到式(4-9)中的指数极限的结论。最后一个表达式是标准高斯随机变量的特征函数，则证明完毕。

由于 $\overline{X}$ 具有均值 μ_X 和方差 σ_X^2/k，以上表达式正好使 $\overline{X}[k]$变换为具有单位方差的零均值随机变量。均值显然为零，方差为：

$$\frac{\mathcal{E}[(\overline{X}[k]-\mu_X)^2]}{\sigma_X^2/k}=\frac{\sigma_X^2/k}{\sigma_X^2/k}=1 \tag{7-62}$$

当然重要的是无论初始分布如何，只要均值存在并且具有有限方差，极限分布就为高斯的。即使没有以上的证明，我们由第 4 章也可以得到独立随机变量的和的 pdf 为各独立 pdf 的卷积。在那里我们给出了连续随机变量的一些例子，证明了随着反复的卷积运算，和值的 pdf 变得越来越平滑，并且开始具有高斯型的样子。对离散随机变量也有相同的结论，例 7-7 对离散的均匀分布随机变量给出了证明。第 4 章描述的离散随机变量的高斯近似是 CLT 的特例。

回顾第 6 章的内容：如果随机变量 X 能被写为 k 个 iid 随机变量的和(对每个 $k\in\mathcal{N}$)，而且所有随机变量具有与 X 相同的分布，则随机变量 X 是*无限可分的*。例子中包括泊松，伽马和学生氏 t 分布。CLT 中用到的样本均值与一个无限可分的随机变量的和是不同的，它因 μ_X 产生了偏移，乘了比值因子$\sqrt{k}/\sigma_X$。这个正态化过程使得式子具有零均值和单位方差，这就是为什么 $Y[k]$依分布收敛于一个标准高斯随机变量的原因。一个无限可分的随机变量不做这种正态化的假设，因此求和项与它的各项可能具有相同的分布。

例 7-7 假设 $X[k]$具有离散的均匀分布，输出为{1, 2}，均值 $\mu_X=(1+2)/2=1.5$，方差 $\sigma_X^2=(2^2-1)/12=1/4$ 则式(7-55)为：

$$Y[k]=\frac{(1/k)\sum_{m=1}^{k}X[m]-1.5}{\sqrt{1/4k}}=\frac{1}{\sqrt{k/4}}\left[\sum_{m=1}^{k}X[m]-1.5k\right] \tag{7-63}$$

对于任意 $k\geqslant 1$，它具有零均值和单位方差。即使 $X[k]$原来只有正整数输出，以上表达式允许负与非整数输出。对求和项，pmf 可由连续的离散卷积得到，而且只对整数值是非零的。上面的偏移和比例运算产生了实数，在表 7-2 中明确的给出了前四个 k 对应的值。显然 k 必须在求和项的 pdf 的尾部开始近似于高斯 pdf 的尾部之前达到足够大。◀

表 7-2 例 7-7 中 CLT 的说明

k	求和项输出	偏移/缩放后输出	概率质量
2	2,3,4	−1.414,0,1.414	1/4,2/4,1/4
3	3,4,5,6	−1.732,−0.577,0.577,1.732	1/8,3/8,3/8,1/8
4	4,5,6,7,8	−2,−1,0,1,2	1/16,4/16,6/16,4/16,1/16
5	5,6,7,8,9,10	−2.236,−1.342,−0.447,0.447,1.342,2.236	1/32,5/32,10/32,10/32,5/32,1/32

例 7-8 设 $X[k]$为一个 iid 指数序列，参数为 λ，均值 $\mu_X=1/\lambda$，方差 $\sigma_X^2=1/\lambda^2$。对这个例子，

$$Y[k]=\frac{(1/k)\sum_{m=1}^{k}X[m]-1/\lambda}{(1/\lambda)/\sqrt{k}}=(\lambda/\sqrt{k})\sum_{m=1}^{k}X[m]-\sqrt{k} \tag{7-64}$$

由于求和项中随机变量都是 iid 的，

$$\Phi_{\sum_{m=1}^{k}X[m]}(\mathrm{j}\omega)=\prod_{m=1}^{k}\Phi_{X[m]}(\omega)=[\lambda/(\lambda-\mathrm{j}\omega)]^k \tag{7-65}$$

它是一个参数为 $\{\lambda, r=k\}$ 埃尔朗（伽马）分布的特征函数。由于求和项乘以了比例因子 $\lambda/\sqrt{k}$（即除以它的标准差），参数变为了 $\{\sqrt{k}, r=k\}$。此时分布的均值和方差分别为 $k/\sqrt{k}=\sqrt{k}$ 和 $k/(\sqrt{k})^2=1$，因此式(7-64)中的 $Y[k]$ 具有零均值和单位方差。图 7-6 中给出了对应不同 k 值时 $Y[k]$ 的 pdf 的例子，为了更好的看清它们的形状将其向右偏移了 $\sqrt{k}$。同时给出均值为 $\mu=\sqrt{20}\approx 4.4721$、具有单位方差的高斯随机变量的 pdf 作为对比。

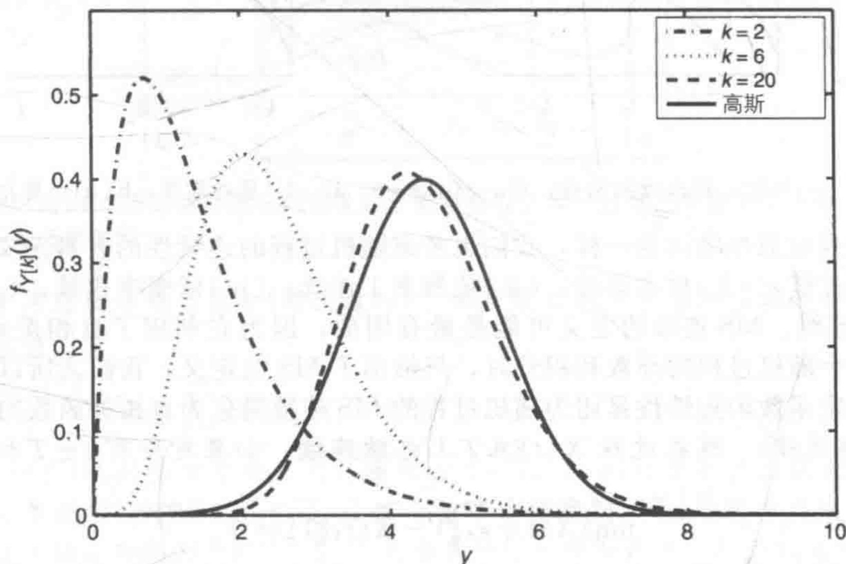


图 7-6　例 7-8 中的 $Y[k]$ 的移位 pdf（为了看起来方便，将其向右偏移了 $\sqrt{k}$）◀

例 7-9　为了应用 CLT，我们要从问题的描述中找出感兴趣的随机变量建模为 iid 随机变量的和。例如，考虑一个学校的综合考试的得分，已知均值为 $\mu_X=70$，标准差 $\sigma_X=10$。则对一个有 10 个人的班，包括样本均值等很多概率都可以用式(7-55)来近似。对 $P(70\leqslant \overline{X}[10]\leqslant 75)$ 中的边界，我们计算出：

$$y_1=\frac{70-70}{10/\sqrt{10}}=0, y_2=\frac{75-70}{10/\sqrt{10}}\approx 1.5811 \tag{7-66}$$

由此得到：

$$\begin{aligned}P(y_1\leqslant Y[10]\leqslant y_2)&=F_Y(1.5811)-F_Y(0)\\&=Q(0)-Q(1.5811)\approx 0.4431\end{aligned} \tag{7-67}$$

其中 $F_Y(x)$ 是标准高斯 cdf，$Q(\cdot)$ 是 Q 函数（注意到 $Q(x)=1-F_Y(x)$）◀

7.5 随机连续

在接下来的几节中，我们考察连续时间随机过程的特性，这些内容在微积分课程中学习过，涉及连续、导数和积分。由于过程是随机的，我们可以检测每个现实的连续性或计算导数和积分，这些都取决于应用的需要。当然，对所有的现实这样做是不可行的，即使可能，我们也不清楚这样的特性描述是否能够提供现实总体集合行为的有用信息。就如在前面的章节中对随机变量和随机过程所做的，我们研究这些微积分的概率公式可能更好。DE 作为线性系统的代表，在输入随机信号时已经讨论过，在第 8 章中还会继续讨论。我们从一个非随机函数的连续性的定义开始。

定义（连续函数）　函数 $x(t)$ 在 t_0 处连续，如果对 $\delta>0$，存在 $\varepsilon>0$，满足

$$|t-t_0|<\varepsilon \Rightarrow |x(t)-x(t_0)|<\delta \tag{7-68}$$

通常，δ 依赖于 ε 和 t_0，我们可以写成函数 $\delta(\varepsilon, t_0)$。这个定义与我们对连续性的直觉是一致的：函数的变量的微小变化不会使函数值发生跃变。图 7-7 描述了一个任意连续函

数的这种情况。其他类型的非随机函数的连续在附录 B 中讨论，并且给出了例子。

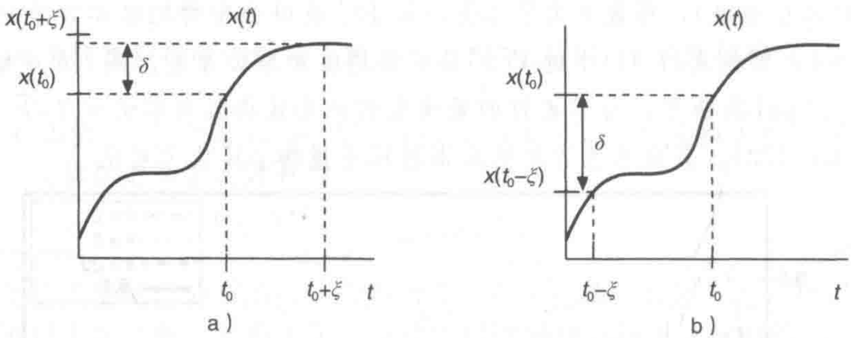


图 7-7 一个在 t_0 处连续的函数：当 $\varepsilon\to 0$，$\delta\to 0$。a) $x(t)$是右连续；b) $x(t)$是左连续

与随机序列收敛性的讨论一样，我们也考虑随机过程的连续性的各种定义。它们分别是：（ⅰ）必然连续；（ⅱ）样本连续；（ⅲ）依概率 1 连续；（ⅳ）依概率连续；（ⅴ）依分布连续；（ⅵ）MS 连续。MS 连续的定义可能是最有用的，因为它利用了自相关函数 $R_{XX}(t_1, t_2)$。当考虑一个随机过程的导数和积分时，只给出了 MS 的定义。我们之所以在附录 B 中详细地讨论确定函数的连续性是因为随机过程的 MS 连续简化为自相关函数的连续。

定义（必然连续） 随机过程 $X(t)$在$\mathcal{T}$上**必然连续**，如果对所有 $t\in\mathcal{T}$和每个 $\zeta\in\mathcal{S}$，满足

$$\lim_{\varepsilon\to 0}|X(t+\varepsilon,\zeta)-X(t,\zeta)|=0 \tag{7-69}$$

由于ζ代表了$X(t)$的一个特定的现实，必然连续主要是针对确定性函数连续性的定义，应用于$\mathcal{S}$中的每个现实。必然连续意味着$\mathcal{T}$上的每个现实都是连续的。这是随机过程连续性定义中最强的，尽管在实际中它不是很有用。例如，如果一个现实碰巧在某时刻 t_0 不连续，这个随机过程就不是必然连续的，即使这个单一的现实不代表过程的总体，它甚至有可能是零概率。用一个概率公式导出下一个连续性的定义。

定义（样本连续） 随机过程 $X(t)$在$\mathcal{T}$上是**样本连续**的，如果

$$P(\lim_{\xi\to 0}|X(t+\varepsilon)-X(t)|\neq 0,\quad 对所有\ t\in\mathcal{T})=0 \tag{7-70}$$

它也可以写作

$$P(X(t)\ 在\ t\ 时刻是连续的(对所有\ t\in\mathcal{T}))=1 \tag{7-71}$$

一个与式(7-69)相似的可选的表达式是

$$P(\zeta\in S:\lim_{\xi\to 0}X(t+\varepsilon,\zeta)-X(t,\zeta)|\neq 0,\quad 对所有\ t\in\mathcal{T})=1 \tag{7-72}$$

在这个定义中所有在区间$\mathcal{T}$上不连续的现实(都具有零概率)都被忽略了；随机过程仍认为是连续的。一个样本连续的随机过程也被称为几乎必然连续。但是，很重要的一点是注意到这个定义针对的现实是在整个区间$\mathcal{T}$上的。接下来连续的定义针对的是现实的总体在一个特定时刻 $t\in\mathcal{T}$。

定义（依概率 1 连续） 随机过程 $X(t)$在 $t\in\mathcal{T}$时刻**依概率 1 连续**，如果：

$$P(\lim_{\varepsilon\to 0}|X(t+\varepsilon)-X(t)|=0)=1 \tag{7-73}$$

它可被写为：

$$P(X(t)\ 在\ t\ 时刻连续)=1 \tag{7-74}$$

另一个与(7-69)相似的表达式是

$$P(\zeta\in\mathcal{S}:\lim_{\xi\to 0}|X(t+\varepsilon,\zeta)-X(t,\zeta)|=0)=1 \tag{7-75}$$

这个连续性的定义意味着$\lim\limits_{\xi\to 0}|X(t+\varepsilon)-X(t)|=0$ 在特定 $t\in\mathcal{T}$时刻发生是几乎必然的。图 7-8 显示的是依概率 1 连续的过程的现实的例子，其中现实在$\mathcal{T}$的不同时刻不连续。如果对任意 $t\in\mathcal{T}$，可数的总概率为零的现实具有不连续性，$X(t)$依概率 1 连续。另一方

面，X(t)不是样本连续的：每个现实在某个 $t\in\mathcal{T}$时刻不连续。以下的例子进一步说明了这个区别。

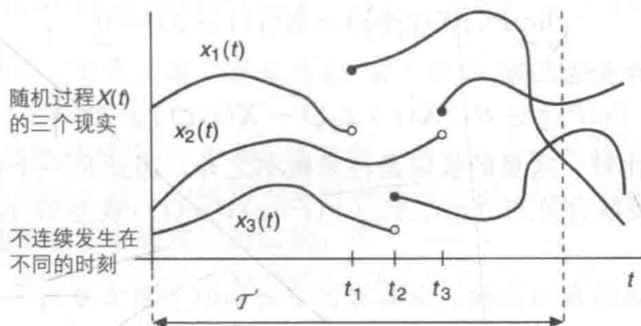


图 7-8 在不同时间 $t\in\mathcal{T}$不连续的现实的例证：它们不是左连续。过程是依概率 1 连续，但不是样本连续，不是必然连续

例 7-10 考虑数字通信信号(随机过程)：大家熟知的二元脉冲振幅调制信号(PAM)，它可以被表示为：

$$X(t)=\sum_{k=0}^{\infty}B[k][u(t-kT_{\circ})-u(t-(k+1)T_{\circ}-T_{\mathrm{d}})] \tag{7-76}$$

其中 $B[k]$是一个 iid 对称伯努利序列，输出为 $\{-1,1\}$，$u(t)$是单位阶跃函数，则$u(t)-u(t-T_{\mathrm{d}})$是一个矩形“脉冲”，$T_{\mathrm{d}}<T_{\circ}$ 是每个脉冲的持续期，$T_{\circ}$ 是脉冲之间的周期。每个现实是一个脉冲波，振幅为±1，只在 $T_{\circ}$ 的整数倍时出现不连续。图 7-9a 显示了 $T_{\mathrm{d}}=T_{\circ}/2$(被称为 50%工作周期，因为脉冲在脉冲周期的一半时回到 0 值)时的两个现实的例子。观察到当相邻的随机变量 $B[k]$和 $B[k-1]$呈现不同值时，不连续发生。显然没有现实是连续的，因此 $X(t)$不是一个样本连续过程。而且，由于不连续发生在相同时刻 $T_{\circ}$，$T_{\circ}+T_{\mathrm{d}}$，$2T_{\circ}$，$2T_{\circ}+T_{\mathrm{d}}$ 等，$X(t)$也不是依概率 1 连续。但是，假设每个实现有一个随机相移 Φ，则随机过程为：

$$X(t)=\sum_{k=0}^{\infty}B[k][u(t-kT_{\circ}-\Phi)-u(t-(k+1)T_{\circ}-T_{\mathrm{d}}-\Phi)] \tag{7-77}$$

假设 Φ 在$[0,\ T_{\circ}-T_{\mathrm{d}})$上均匀分布，则任何现实中的脉冲不会延伸到相邻时间间隔。在一个特定时刻 t 发生不连续的概率为零，因为 Φ 是一个连续随机变量。图 7-9b 中显示了具有随机相移的两个现实的例子。即使所有的现实仍不是样本连续的，具有随机相移的 $X(t)$是依概率 1 连续的，它是对每个 $t\in\mathcal{T}$的总体定义的，不是针对间隔持续期。

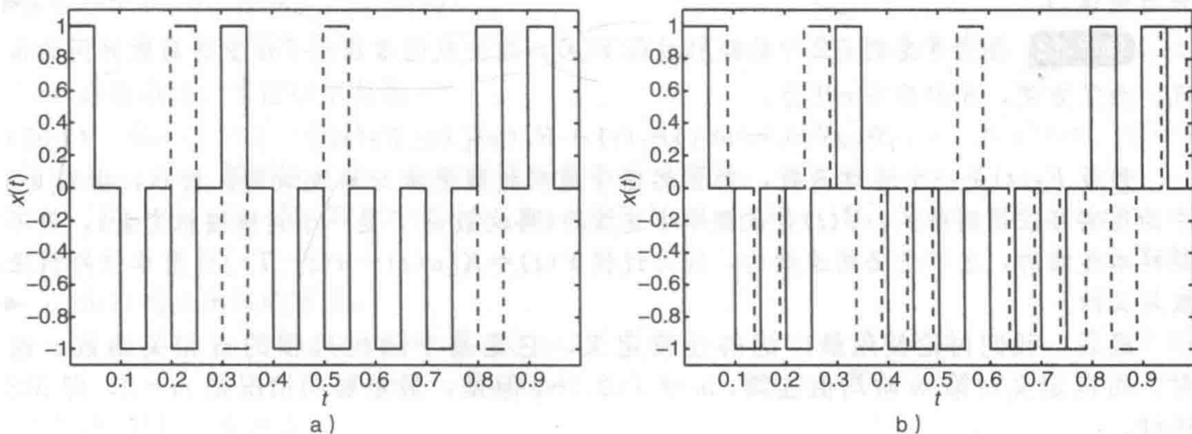


图 7-9 二元 PAM 的现实的例子。a)相移为零的二元 PAM 波形；b)具有随机相移的二元 PAM 波形(当没有相移，且两个波形的脉冲相同时，实线部分掩盖了虚线) ◀

剩下的连续的定义不考虑各单独的现实，关注的是过程的分布或矩。

定义(依概率连续)　随机过程 $X(t)$在 $t\in\mathcal{T}$时刻**依概率连续**，如果对每个 $\delta>0$

$$\lim_{\varepsilon\to 0}P(|X(t+\varepsilon)-X(t)|\geqslant\delta)=0 \tag{7-78}$$

与式(7-75)类似的表达式为：

$$\lim_{\varepsilon\to 0}P(\zeta\in\mathcal{S}:|X(t+\varepsilon,\zeta)-X(t,\zeta)|\geqslant\delta)=0 \tag{7-79}$$

与依概率1连续比较，这里的极限是在求概率之外，因此是一个稍弱的定义。为了评价式(7-78)，我们需要联合分布 $f_{X(t+\varepsilon),X(t)}(x(t+\varepsilon),x(t))$或差值 $f_{X(t+\varepsilon)-X(t)}(x(t+\varepsilon)-x(t))$的 pdf。

例 7-11　考虑扩展的维纳过程，它具有以下的 pdf(见第6章)：

$$f_{X(t_2)-X(t_1)}(x)=\frac{1}{\sqrt{2\pi(t_2-t_1)}}\exp(-x^2/2(t_2-t_1)) \tag{7-80}$$

代入 $t_2=t+\varepsilon$ 和 $t_1=t$，式(7-78)变为

$$P(|X(t+\varepsilon)-X(t)|\geqslant\delta)=2\int_\delta^\infty\frac{1}{\sqrt{2\pi\varepsilon}}\exp(-x^2/2\varepsilon)\mathrm{d}x=2Q(\delta/\sqrt{\varepsilon}) \tag{7-81}$$

其中我们用到了 pdf 的对称性，$Q(\cdot)$是 Q 函数。扩展的维纳过程是依概率连续的，因为对每个 $\delta>0$ 有$\lim_{\varepsilon\to 0}Q(\delta/\sqrt{\varepsilon})=0$。◀

最弱的连续类型是基于 cdf 的。

定义(依分布连续)　随机过程 $X(t)$在 $t\in\mathcal{T}$时**依分布连续**，如果满足

$$\lim_{\varepsilon\to 0}F_{X(t+\varepsilon)}(x)=F_{X(t)}(x) \tag{7-82}$$

对每个 $x\in\mathcal{R}$成立。

这个极限只是指 cdf 本身的连续性。如果过程是 iid 的，显然它是依分布连续的，因为 cdf 独立于 t。

例 7-12　考虑标准维纳过程，其 cdf 为：

$$F_{X(t+\varepsilon)}(x)=\int_x^\infty\frac{1}{\sqrt{2\pi(t+\varepsilon)}}\exp(-z^2/2(t+\varepsilon))\mathrm{d}z=Q(x/\sqrt{t+\xi}) \tag{7-83}$$

由于 Q 函数是连续的，

$$\lim_{\varepsilon\to 0}Q(x/\sqrt{t+\varepsilon})=Q(x/\sqrt{t}) \tag{7-84}$$

则维纳过程是依分布连续的。(这个结果也可从例 7-11 中得到，因为依概率连续意味着依分布连续。)◀

例 7-13　再次考虑例 6-2 中的随机过程 $Y(t)$，这次我们考虑例 7-6 中的离散时间的情况。为了方便，我们重写 cdf 为：

$$F_{Y(t)}(y)=u(y)F_T(t)+F_X(y)[1-F_T(t)] \tag{7-85}$$

假设 $F_T(t)$是一个连续函数，则显然这个随机过程是依分布连续的。而且，由图 6-5 中的现实可以得到结论：$Y(t)$是依概率1连续的(再次假设 T 是一个连续随机变量)，但不是样本连续的，也不是必然连续的，因为过程 $Y(t)=X[u(t)-u(t-T)]$是用单位阶跃函数定义的。◀

最后，我们讨论使用最广泛的连续定义，它是基于随机过程的自相关函数。通常，可以定义对第 m 阶均值连续，$m=1,2,\cdots$。但是，最重要的情况是 $m=2$，即 MS 连续。

定义(均方连续)　随机过程 $X(t)$在 $t\in\mathcal{T}$是 **MS 连续**的，如果$\mathcal{E}[X^2(t)]<\infty$并且

$$\lim_{\varepsilon\to 0}\mathcal{E}[(X(t+\varepsilon)-X(t))^2]=0 \tag{7-86}$$

例 7-14 由扩展维纳过程，同时它也是一个独立增量过程，我们得到：

$$\mathcal{E}[(X(t+\varepsilon)-X(t))^2]=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\varepsilon}}x^2\exp(-x^2/2\varepsilon)\mathrm{d}x=\varepsilon \tag{7-87}$$

因为增量过程是高斯的，均值为零，方差为 ε。由于表达式在 $\varepsilon\to 0$ 时趋于零，扩展维纳过程是 MS 连续的。◀

定理 7-8 如果随机过程 $X(t)$ 是 MS 连续的，则它也是依均值连续的。

证明： 由于一个随机变量的均值平方不可能超过它的 MS，同样的一个随机过程在两个不同时刻对应的随机变量也有相同的结论，即

$$\mathcal{E}[(X(t+\varepsilon)-X(t))^2]\geqslant\mathcal{E}^2[X(t+\varepsilon)-X(t)] \tag{7-88}$$

则

$$\begin{aligned}\lim_{\varepsilon\to 0}\mathcal{E}((X(t+\varepsilon)-X(t))^2]=0&\Rightarrow\lim_{\varepsilon\to 0}\mathcal{E}^2[X(t+\varepsilon)-X(t)]=0\\&\Rightarrow\lim_{\varepsilon\to 0}\mathcal{E}[X(t+\varepsilon)-X(t)]=0\\&\Rightarrow\lim_{\varepsilon\to 0}\mathcal{E}[X(t+\varepsilon)]=\mathcal{E}[X(t)]\end{aligned} \tag{7-89}$$

证明完毕。

依均值连续并不意味着 MS 连续。如果均值 $\mu_X(t)$ 不是连续函数，则 $X(t)$ 不是 MS 连续的，因为以下的下边界：

$$\mathcal{E}[(X(t+\varepsilon)-X(t))^2]\geqslant(\mathcal{E}[X(t+\varepsilon)]-\mathcal{E}[X(t)])^2 \tag{7-90}$$

如果对 $t\in\boldsymbol{T}$，当 $\varepsilon\to 0$ 时，不等式的右边不趋向于零，则式(7-86)中 MS 连续的条件无法达到。

观察到式(7-86)中的二阶矩可以写成自相关函数的形式，如下所示：

$$\begin{aligned}\mathcal{E}[(X(t+\varepsilon)-X(t))^2]=&R_{XX}(t+\varepsilon,t+\varepsilon)+R_{XX}(t,t)\\&-R_{XX}(t+\varepsilon,t)-R_{XX}(t,t+\varepsilon)\end{aligned} \tag{7-91}$$

由这个表达式，我们得到如果自相关函数对两个自变量都是连续的，则可以得到 MS 连续。我们只需检查 $R_{XX}(t_1,t_2)$ 在接近直线 $t_1=t_2\in\boldsymbol{T}$ 时的连续性，而不需要对每个 $\{t_1,t_2\}$ 进行分析。

定理 7-9 $\mathcal{E}[X^2(t)]<\infty$ 的随机过程 $X(t)$ 在 $t\in\boldsymbol{T}$ 处是 MS 连续的，当且仅当 $R_{XX}(t_1,t_2)$ 在 $t_1=t_2=t$ 处连续。

证明： (Larson 和 Shubert，1979)充分条件：由式(7-91)，如果 $R_{XX}(t_1,t_2)$ 在 t 时刻对两个变量都是连续的，则等式右边 $\to 0$。由具有两个变量函数的连续性的定义，这意味着对每个 $\delta>0$，存在 $\varepsilon>0$，使得：

$$\sqrt{(t_1-t)^2+(t_2-t)^2}<\varepsilon\Rightarrow|R_{XX}(t_1,t_2)-R_{XX}(t,t)|<\delta \tag{7-92}$$

必要条件：考虑以下差值

$$R_{XX}(t_1,t_2)-R_{XX}(t,t)=R_{XX}(t_1,t_2)-R_{XX}(t,t_1)+R_{XX}(t,t_1)-R_{XX}(t,t) \tag{7-93}$$

其中 $R_{XX}(t,t_1)$ 被引入式中。由绝对值三角不等式得到：

$$|R_{XX}(t_1,t_2)-R_{XX}(t,t)|\leqslant|R_{XX}(t_1,t_2)-R_{XX}(t,t_1)|+|R_{XX}(t,t_1)-R_{XX}(t,t)| \tag{7-94}$$

由自相关函数的定义，

$$R_{XX}(t_1,t_2)-R_{XX}(t_1,t)=\mathcal{E}[X(t_1)(X(t_2)-X(t))] \tag{7-95}$$

$$R_{XX}(t,t_1)-R_{XX}(t,t)=\mathcal{E}[X(t)(X(t_1)-X(t))] \tag{7-96}$$

用柯西-施瓦茨不等式：

$$|R_{XX}(t_1,t_2)-R_{XX}(t_1,t)|\leqslant\sqrt{\mathcal{E}[X^2(t_1)]\,\mathcal{E}[(X(t_2)-X(t))^2]} \tag{7-97}$$

$$|R_{XX}(t,t_1)-R_{XX}(t,t)|\leqslant\sqrt{\mathcal{E}[X^2(t)]\,\mathcal{E}[(X(t_1)-X(t))^2]} \tag{7-98}$$

代入 $t_1=t+\varepsilon_1$ 和 $t_2=t+\varepsilon_2$，用式(7-86)中的 MS 连续的定义，得到当 ε_1，$\varepsilon_2\to 0$ 时，式(7-94)的右边趋于零，证明完毕。

对 MS 连续，我们只对 $R_{XX}(t_1, t_2)$在 $t_1=t_2=t$ 处连续感兴趣。我们需要从任意$\{t_1, t_2\}\in\mathcal{T}$开始评价连续性，通过检查 $R_{XX}(t_1, t_2)$，其中 $t_1=t+\varepsilon_1$，$t_2=t+\varepsilon_2$，令 ε_1 和 ε_2 各自趋于零。这确保了连续性不是仅沿着直线 $t_1=t_2$ 来评估，避免了只检测当 $\varepsilon\to 0$ 时 $R_{XX}(t+\varepsilon, t+\varepsilon)$ 的错误行为(尽管通常这是一个更容易检测的，可用来确定一个随机过程不是 MS 连续的首要条件)。最后，如果 $X(t)$对每个 $t\in\mathcal{T}$是 MS 连续的。则 $R_{XX}(t_1,t_2)$ 对所有 t_1，$t_2\in\mathcal{T}$都是连续的(见习题 7-18)。

例 7-15 泊松计数过程具有以下的自相关函数：

$$R_{XX}(t_1,t_2)=\lambda\min(t_1,t_2)+\lambda^2 t_1 t_2 \tag{7-99}$$

其中 $\lambda>0$ 是比值参数。由定理 7-9：

$$\begin{aligned}&\lim_{\varepsilon_1,\varepsilon_2\to 0}|R_{XX}(t+\varepsilon_1,t+\varepsilon_2)-R_{XX}(t,t)|\\&=\lim_{\varepsilon_1,\varepsilon_2\to 0}|\lambda\min(t+\varepsilon_1,t+\varepsilon_2)+\lambda^2(t+\varepsilon_1)(t+\varepsilon_2)-\lambda t-\lambda^2 t^2|\\&=\lim_{\varepsilon_1,\varepsilon_2\to 0}|\lambda\min(\varepsilon_1,\varepsilon_2)+\lambda^2(t\varepsilon_1+t\varepsilon_2+\varepsilon_1\varepsilon_2)|=0\end{aligned} \tag{7-100}$$

即使现实不是样本连续的，泊松过程是 MS 连续的，如图 7-10 所示。图 6-20 是其自相关函数的曲线，由图我们看到 $R_{XX}(t_1,t_2)$ 在 $t_1=t_2\in\mathcal{T}$处是连续的。

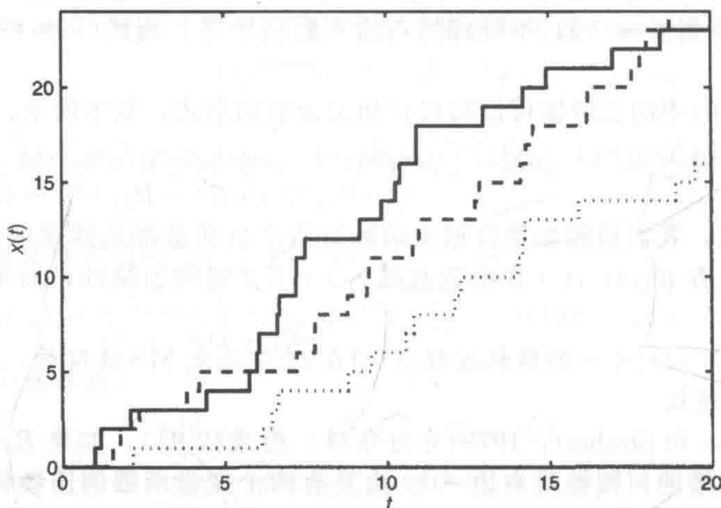


图 7-10 泊松计数过程的现实的例子，显示它们各自都不是样本连续的，尽管过程整体是 MS 连续的 ◀

例 7-16 “半随机”电报信号 $Y(t)$的现实由泊松计数过程 $X(t)$按如下方法产生(见第 6 章)：$Y(t)=(-1)^{X(t)}$。随着泊松过程的每次跃变，随机电报信号的符号都会改变，保持对增量数目的奇、偶校验。如果 $Y(t)$乘以一个输出等概率取$\{-1,1\}$的对称伯努利随机变量，且这个随机变量独立于$Y(t)$，则得到了电报信号的“随机化”形式。这引入了不确定性(更大的随机性)，因为我们不再知道-1和 1 哪个对应增量的偶数值。图 7-11 给出了现实的例子。随机电报信号的两种形式都具有以下的自相关函数(见习题 6-3 和表 6-4)

$$R_{YY}(t_1,t_2)=\exp(-2\lambda|t_2-t_1|) \tag{7-101}$$

在图 7-12a 中给出了。这个过程是 MS 连续的：

$$\lim_{\varepsilon_1,\varepsilon_2\to 0}|R_{YY}(t+\varepsilon_1,t+\varepsilon_2)-R_{YY}(t,t)|=\lim_{\varepsilon_1,\varepsilon_2\to 0}|\exp(-2\lambda|\varepsilon_1-\varepsilon_2|)-1|=0 \tag{7-102}$$

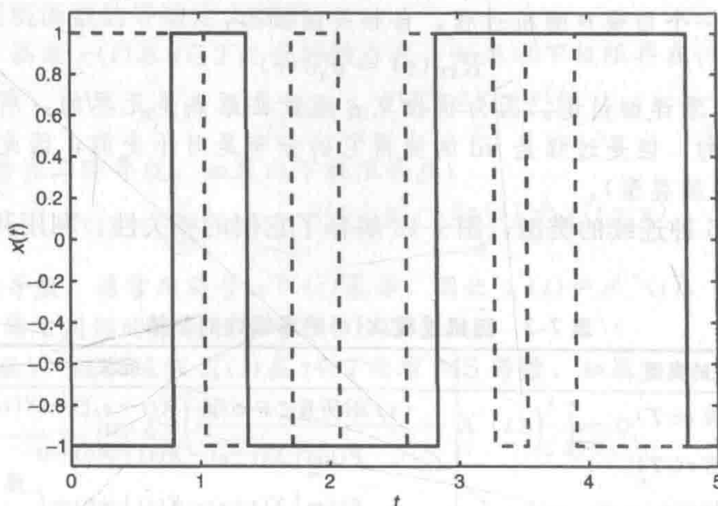


图 7-11　随机电报信号的现实，显示它们各自都不是样本连续的，尽管过程是 MS 连续的

随机化的电报信号实际上是广义平稳的，因为可用 $\tau = t_2 - t_1$ 来替换。自相关函数 $R_{YY}(\tau)$ 只依赖于一个变量，而且它的均值为零。尽管半随机电报信号的自相关函数也只依赖于时间间隔 τ，它并不是广义平稳的，因为它的均值随时间而改变。对 $R_{XX}(\tau)$，MS 连续变为只在 $\tau=0$ 这一点连续。

定理 7-10　如果随机过程 $X(t)$ 是广义平稳的，则当且仅当 $R_{YY}(\tau)$ 在 $\tau=0$ 处连续，它是 MS 连续的。

证明：这是定理 7-9 的特例，式(7-91)变为：

$$\mathcal{E}[(X(t+\varepsilon)-X(t))^2] \leqslant R_{XX}(0) - R_{XX}(\varepsilon) + R_{XX}(0) - R_{XX}(\varepsilon) \tag{7-103}$$

证明过程基本上相同(还更简单些，因为 $R_{XX}(\tau)$ 是只有一个变量的函数)。

对广义平稳的情况，我们需要检测 $R_{YY}(\tau)$ 在 $\tau=0$ 处的连续性，可从两个方向 $\tau\pm\varepsilon$ 逼近(不同于 $R_{XX}(t_1,t_2)$，它的连续性需要在 t_1-t_2 平面上从各个方向来检测在 $t_1=t_2=t$ 处的连续性)。

例 7-17　对例 7-16 中的随机电报信号，自相关函数可以写为：

$$R_{XX}(\tau) = \exp(-2\lambda|\tau|) \tag{7-104}$$

它在原点处连续：

$$\lim_{\tau\to 0}|R_{XX}(\tau) - R_{XX}(0)| = \lim_{\tau\to 0}|\exp(-2\lambda|\tau|) - 1| = 0 \tag{7-105}$$

图 7-12b 给出了 $R_{YY}(\tau)$ 曲线，它看起来相当于一个“薄片”沿着任意一条直线穿过 $R_{XX}(t_1,t_2)$ 垂直于图 7-12a 中的 $t_1=t_2$ 直线。广义平稳过程的自相关函数必须具有对 $\tau=t_2-t_1$ 的对称性。

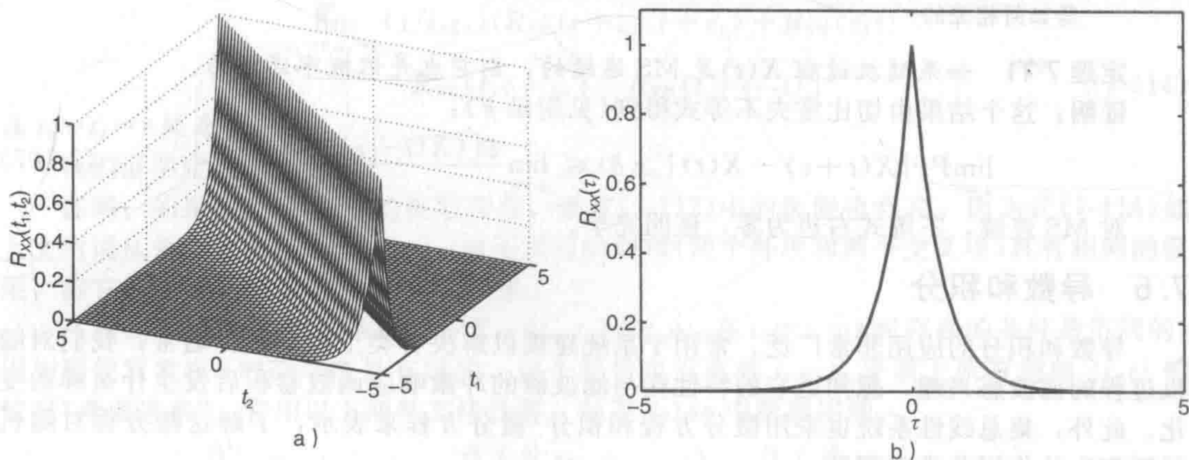


图 7-12　随机电报信号的自相关函数，说明了它在 $t_1=t_2$ 处连续。a) $R_{XX}(t_1,t_2)$ 在 t_1-t_2 平面；b) $R_{YY}(\tau)$ 与 τ 的关系。由于 $X(t)$ 是广义平稳的，$R_{YY}(\tau)$ 只需在 $\tau=0$ 处检测它的连续性

例 7-18 考虑一个白噪声随机过程，自相关函数为：

$$R_{XX}(\tau)=\sigma_X^2\delta(\tau) \tag{7-106}$$

（白噪声在第 8 章详细讨论。）因为狄拉克 δ 函数在原点是无界的，所以所有的这类过程都不是 MS 连续的，但是过程是 iid 的使得它的分布是时不变的，因此是依分布连续的（这是一种较弱的连续类型）。◀

表 7-3 总结了 6 种连续的类型，图 7-13 解释了它们的相关性。利用其中一个的关系证明来总结这一节。

表 7-3　随机过程 $X(t)$ 的连续性的总结

连续性的类型	定义
必然连续（对所有 $t\in T$）	对所有 $\zeta\in S$ $\lim\limits_{\varepsilon\to 0}\lvert X(t+\varepsilon,\zeta)-X(t,\zeta)\rvert=0$
样本连续（对所有 $t\in T$）	$P(\lim\limits_{\varepsilon\to 0}\lvert X(t+\varepsilon)-X(t)\rvert\neq 0)=0$
依概率 1 连续（对所有 $t\in T$）	$P(\lim\limits_{\varepsilon\to 0}\lvert X(t+\varepsilon)-X(t)\rvert=0)=1$
依概率连续（在 $t\in T$）	对每个 $\delta>0$ $\lim\limits_{\varepsilon\to 0}P(\lvert X(t+\varepsilon)-X(t)\rvert\geqslant\delta)=0$
均方连续（在 $t\in T$）	$\lim\limits_{\varepsilon\to 0}\mathcal{E}[\lvert X(t+\varepsilon)-X(t)\rvert^2]=0$
依分布连续（在 $t\in T$）	$\lim\limits_{\varepsilon\to 0}F_{X(t+\varepsilon)}(x)=F_{X(t)}(x)$
均方连续	**条件**
非平稳（在 $t\in T$）	$R_{XX}(t_1,t_2)$ 在 $t_1=t_2=t$ 处连续
广义平稳（在 $t\in T$）	$R_{XX}(\tau)$在 $\tau=0$ 处连续

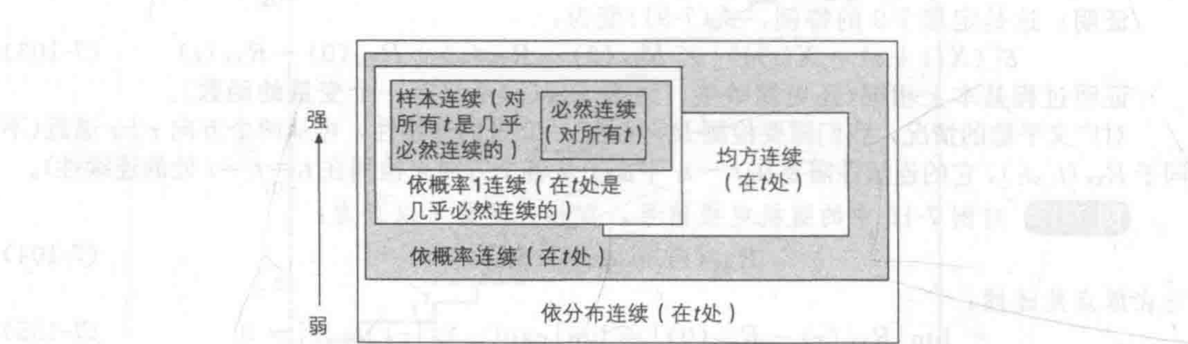


图 7-13　随机过程的六种连续之间的关系。必然连续⊆样本连续⊆依概率 1 连续⊆依概率连续⊆依分布连续，$L^2\subseteq p\subseteq d$。通常，我们不能推断出 MS 连续与依概率 1 连续、样本连续和必然连续是如何相关的

定理 7-11　如果随机过程 $X(t)$是 MS 连续的，则它也是依概率连续的。

证明：这个结果由切比雪夫不等式得到(见附录 F)：

$$\lim_{\varepsilon\to 0}P(\lvert X(t+\varepsilon)-X(t)\rvert\geqslant\delta)\leqslant\lim_{\varepsilon\to 0}\frac{\mathcal{E}[(X(t+\varepsilon)-X(t))^2]}{\delta^2} \tag{7-107}$$

对 MS 连续，不等式右边为零，证明完毕。

7.6　导数和积分

导数和积分的应用非常广泛，常用于系统建模以解决各类工程问题。通常，我们对随机过程的滤波感兴趣，想知道它的特性在与滤波器的冲激响应函数卷积后发生什么样的变化。此外，集总线性系统也采用微分方程和积分-微分方程来表示，了解这些方程对随机过程产生的作用非常有帮助。

我们只考虑随机过程的导数和积分的 MS 定义，它会再次导出自相关函数 $R_{XX}(t_1,t_2)$

的评估。一个非随机函数的导数定义如下：

定义(导数)　函数 $x(t)$ 在 $t\in\mathcal{T}$ 处有**导数**存在，如果以下极限存在：

$$x'(t)=\frac{\mathrm{d}x}{\mathrm{d}t}\triangleq\lim_{\varepsilon\to0}\frac{x(t+\varepsilon)-x(t)}{\varepsilon} \tag{7-108}$$

它在 $t\in\mathcal{T}$ 处存在二阶导数，如果以下极限存在：

$$x''(t)=\frac{\mathrm{d}^2x}{\mathrm{d}t^2}\triangleq\lim_{\varepsilon\to0}\frac{x(t+\varepsilon)-2x(t)+x(t-\varepsilon)}{\varepsilon^2} \tag{7-109}$$

对更高阶一般导数，通常用符号 $x^{(m)}(t)$ 表示，因此 $x'(t)=x^{(1)}(t)$，$x''(t)=x^{(2)}(t)$。将这个定义与 MS 连续中用到的方法联合可导出 MS 导数。

定义(均方导数)　随机过程 $X(t)$ 在 $t\in\mathcal{T}$ 处有 MS **导数**，如果

$$\lim_{\varepsilon\to0}\mathcal{E}\left[\left(\frac{X(t+\varepsilon)-X(t)}{\varepsilon}-X'(t)\right)^2\right]=0 \tag{7-110}$$

其中 $X'(t)$ 代表导数。

除了 $X(t)$ 是一个随机过程，式(7-110)中第一项是有限近似值。第二项 $X'(t)$ 代表导数过程，由极限中的有限近似产生。如果 $X'(t)$ 事先已知，则可以评估这个表达式，因为期望值是在求极限前计算的。但是，导数过程通常是不知道的，因此由柯西准则来确定一个随机过程是否具有 MS 导数：

$$\lim_{\varepsilon_1\to0,\varepsilon_2\to0}\mathcal{E}\left[\left(\frac{X(t+\varepsilon_1)-X(t)}{\varepsilon_1}-\frac{X(t+\varepsilon_2)-X(t)}{\varepsilon_2}\right)^2\right]=0 \tag{7-111}$$

考虑表达式中以下项：

$$\mathcal{E}\left[\left(\frac{X(t+\varepsilon_1)-X(t)}{\varepsilon_1}\right)^2\right]=(1/\varepsilon_1^2)[R_{XX}(t+\varepsilon_1,t+\varepsilon_1)+R_{XX}(t,t)-R_{XX}(t,t+\varepsilon_1)-R_{XX}(t+\varepsilon_1,t)] \tag{7-112}$$

对包含 ε_2 的项也有同样的结果。对式(7-111)中的两个交叉项有类似的结果：

$$\mathcal{E}\left[\frac{X(t+\varepsilon_1)-X(t)}{\varepsilon_1}\frac{X(t+\varepsilon_2)-X(t)}{\varepsilon_2}\right]=(1/\varepsilon_1\varepsilon_2)[R_{XX}(t+\varepsilon_1,t+\varepsilon_2)+R_{XX}(t,t)-R_{XX}(t,t+\varepsilon_2)-R_{XX}(t+\varepsilon_1,t)] \tag{7-113}$$

尽管这两个表达式类似，但式(7-113)中的交叉项与 ε_1 和 ε_2 相关。这意味着求极限时，当 ε_1 和 ε_2 趋近于零，$R_{XX}(t_1,t_2)$ 导数需要从 t_1-t_2 平面上每个方向来进行评估。式(7-112)中的情况不同，因为每一项仅与 ε_1 或 ε_2 相关(不是与两个同时相关)。

定理 7-12　自相关函数为 $R_{XX}(t_1,t_2)$ 的随机过程 $X(t)$，在 $t\in\mathcal{T}$ 时具有 MS 导数，当且仅当

$$\lim_{\varepsilon_1\to0,\varepsilon_2\to0}(1/\varepsilon_1\varepsilon_2)(R_{XX}(t+\varepsilon_1,t+\varepsilon_2)+R_{XX}(t,t)-R_{XX}(t,t+\varepsilon_2)-R_{XX}(t+\varepsilon_1,t)] \tag{7-114}$$

在 $t_1=t_2=t$ 处存在。

我们证明定理的充分性。

证明：如果式(7-114)中的极限存在，则式(7-112)中的极限也存在，因为式(7-114)如上文所说从每个方向趋近于 $\{t,t\}$。由于所有的四项(两个外项和两个交叉项)具有相同的极限，将它们联合得到式(7-111)中的结果。

注意到式(7-114)中的条件与 $\partial^2R_{XX}(t_1,t_2)/\partial t_1\partial t_2$ 在 $t_1=t_2=t$ 时存在的条件是不同的，因为即使后者在 t 时刻可能是有限的，这个符号也没有将 t_1-t_2 平面上所有趋向 $\{t,t\}$ 的情况“考虑进来”。它用以下两种方法计算，如图 7-14a 中所描述的：

$$\frac{\partial^2}{\partial t_1\partial t_2}R_{XX}(t_1,t_2)=\frac{\partial}{\partial t_1}\left(\frac{\partial}{\partial t_2}R_{XX}(t_1,t_2)\right)=\frac{\partial}{\partial t_2}\left(\frac{\partial}{\partial t_1}R_{XX}(t_1,t_2)\right) \tag{7-115}$$

第一种方法，偏导是平行于 t_2 轴计算的(t_1 固定)，然后沿 t_1 轴计算(t_2 固定)。第二

种方法微分的顺序正好相反。由图中，我们看到导数只需沿着 $\{t,t\}$ 处交叉的两条垂直线检测即可。另一方面，式(7-114)中的条件非常清晰，因为它将 t_1-t_2 平面上所有趋向 $\{t,t\}$ 的情况都考虑了，图 7-14b 中的例子说明了这点。

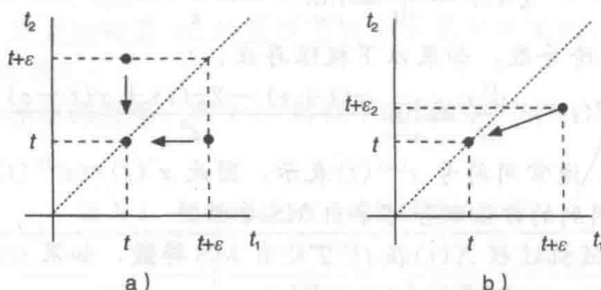


图 7-14 基于 $R_{XX}(t_1,t_2)$ 的 MS 导数。a) 式(7-115)中顺序求偏导的描述；b) 用式(7-114)求 MS 导数的存在性

以下的充分条件通常用于 MS 导数中，我们不加证明的给出。注意到它需要偏导的连续性。

定理 7-13 如果式(7-115)中的偏导数存在并且对每个 $t_1=t_2=t\in\mathcal{T}$ 都是连续的，自相关函数为 $R_{XX}(t_1,t_2)$ 的随机过程 $X(t)$ 在 $t\in\mathcal{T}$ 时具有 MS 导数。

例 7-19 考虑一个泊松过程，自相关函数为式(7-99)，见图 6-20。可以非常方便的应用如下恒等式(见附录 E)

$$\min(t_1,t_2)=(1/2)(t_1+t_2-|t_1-t_2|) \tag{7-116}$$

因此：

$$R_{XX}(t_1,t_2)=(\lambda/2)(t_1+t_2-|t_1-t_2|)+\lambda^2 t_1 t_2 \tag{7-117}$$

微分得到：

$$\frac{\partial}{\partial t_1}R_{XX}(t_1,t_2)=(\lambda/2)\mathrm{sgn}(t_2-t_1)+\lambda^2 t_2 \tag{7-118}$$

$$\frac{\partial^2}{\partial t_1\partial t_2}R_{XX}(t_1,t_2)=\lambda\delta(t_2-t_1)+\lambda^2 \tag{7-119}$$

从中可以看出因为狄拉克 δ 函数使得导数在 $t_1=t_2=t$ 处不存在。式(7-118)中的一阶导数由图 7-15 给出。泊松计数过程没有 MS 导数。

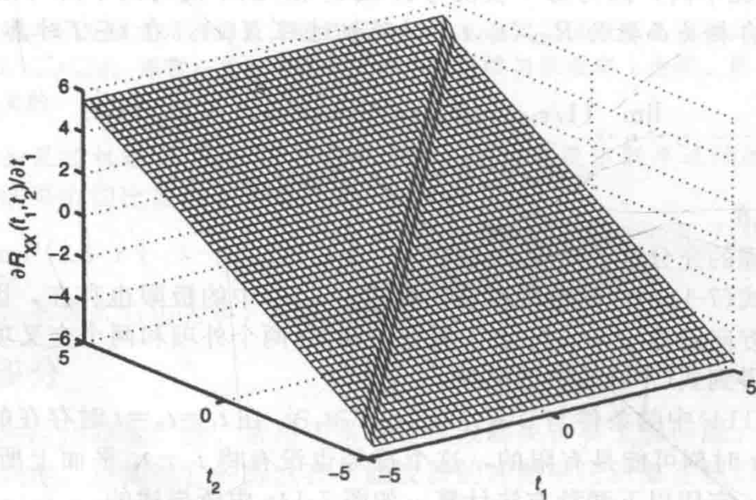


图 7-15 参数 $\lambda=1$ 的泊松计数过程的自相关函数的导数：$\partial R_{XX}(t_1, t_2)/\partial t_1$

对一个广义平稳的过程，只需检测 $\tau=0$ 处的自相关函数 $R_{XX}(\tau)$。

定理 7-14　广义平稳随机过程 $X(t)$ 对每个 $t\in T$ 都有 MS 导数，当且仅当在 $\tau=0$ 处有 $d^2R_{XX}(\tau)/d\tau^2$ 存在。

我们证明定理的充分性条件。

证明：为了证明式(7-111)，广义平稳过程的 $R_{XX}(\tau)$ 在式(7-112)中的外项变为：

$$\mathcal{E}\left[\left(\frac{X(t+\varepsilon_1)-X(t)}{\varepsilon_1}\right)^2\right]=(1/\varepsilon_1^2)[2R_{XX}(0)-2R_{XX}(\varepsilon_1)] \tag{7-120}$$

用式(7-109)中二阶一般导数的定义和 $R_{XX}(\tau)$ 为偶函数的性质，我们得到以下在 $\tau=0$ 处的极限值：

$$\lim_{\varepsilon_1\to 0}(1/\varepsilon_1^2)[2R_{XX}(0)-2R_{XX}(\varepsilon_1)]=-\frac{d^2}{d\tau^2}R_{XX}(\tau)\big|_{\tau=0}=-R''_{XX}(0)$$

类似的，对式(7-113)中交叉项有：

$$\mathcal{E}\left[\left(\frac{X(t+\varepsilon_1)-X(t)}{\varepsilon_1}\right)\left(\frac{X(t+\varepsilon_2)-X(t)}{\varepsilon_2}\right)\right]=(1/\varepsilon_1\varepsilon_2)[R_{XX}(\varepsilon_2-\varepsilon_1)+R_{XX}(0)-R_{XX}(\varepsilon_1)-R_{XX}(\varepsilon_2)] \tag{7-121}$$

在 $\tau=0$ 和 $\tau=\varepsilon_2$ 处将式(7-108)用两次，得到：

$$\lim_{\varepsilon_1\to 0}(1/\varepsilon_1\varepsilon_2)[R_{XX}(\varepsilon_2-\varepsilon_1)-R_{XX}(\varepsilon_2)+R_{XX}(0)-R_{XX}(-\varepsilon_1)]=(1/\varepsilon_2)[-R'_{XX}(\varepsilon_2)+R'_{XX}(0)] \tag{7-122}$$

其中

$$R'_{XX}(0)=\frac{d}{d\tau}R_{XX}(\tau)\big|_{\tau=0} \tag{7-123}$$

则

$$\lim_{\varepsilon_2\to 0}-(1/\varepsilon_2)[R'_{XX}(\varepsilon_2)-R'_{XX}(0)]=-R''_{XX}(0) \tag{7-124}$$

将以上所有 4 个结果联合，得到：

$$\lim_{\varepsilon_1\to 0,\varepsilon_2\to 0}\mathcal{E}\left[\left(\frac{X(t+\varepsilon_1)-X(t)}{\varepsilon_1}-\frac{X(t+\varepsilon_2)-X(t)}{\varepsilon_2}\right)^2\right]=2R''_{XX}(0)-2R''_{XX}(0)=0 \tag{7-125}$$

证明完毕。

与一般的非平稳情况不同，求定理 7-14 中的二阶导数是没有问题的，它给出了充分必要条件。这是因为 $R_{XX}(\tau)$ 是一个只有一个变量 τ 的函数，因此导数只需从 $\tau\pm\xi$ 两个方向检测。最后，注意到由于 $R_{XX}(\tau)$ 是偶函数，它的导数是个奇函数：

$$\frac{d}{d\tau}R_{XX}(\tau)=-\frac{d}{d\tau}R_{XX}(-\tau) \tag{7-126}$$

因此，如果上式存在，则 $R'_{XX}(0)=0$。当然，这并不意味这着二阶导数也为零，因为 $R''_{XX}(\tau)$ 必须在 $\tau=0$ 被代入前计算。

接下来，由于 $X'(t)$ 是一个随机变量，我们给出它在 t 处的均值的表达式：

$$\mathcal{E}[X'(t)]=\mathcal{E}\left[\lim_{\varepsilon\to 0}\frac{X(t+\varepsilon)-X(t)}{\varepsilon}\right]=\lim_{\varepsilon\to 0}(1/\varepsilon)\,\mathcal{E}[X(t+\varepsilon)-X(t)]=\frac{d}{dt}\mathcal{E}[X(t)] \tag{7-127}$$

和自相关函数：

$$\begin{aligned}R_{X'X'}(t_1,t_2)&=\lim_{\varepsilon_1,\varepsilon_2\to 0}\mathcal{E}\left[\frac{(X(t_1+\varepsilon_1)-X(t_1))(X(t_2+\varepsilon_2)-X(t_2))}{\varepsilon_1\varepsilon_2}\right]\\&=\lim_{\varepsilon_1,\varepsilon_2\to 0}(1/\varepsilon_1\varepsilon_2)[R_{XX}(t_1+\varepsilon_1,t_2+\varepsilon_2)+R_{XX}(t_1,t_2)\\&\quad-R_{XX}(t_1+\varepsilon_1,t_2)-R_{XX}(t_1,t_2+\varepsilon_2)],\\&=\frac{\partial^2}{\partial t_1\partial t_2}R_{XX}(t_1,t_2)\end{aligned} \tag{7-128}$$

注意最后一个表达式和定理 7-13 中用来确定 MS 导数是否存在的充分条件的相似性。上文中我们看到导数过程的自相关函数是初始过程 $X(t)$ 的自相关函数的二阶导数，$R_{X'X'}(t_1, t_2)$ 对所有 $\{t_1, t_2\}$ 值都有定义。定理陈述了 $R_{X'X'}(t_1, t_2)$ 的存在以及对所有 $t_1 = t_2 = t \in \mathcal{T}$ 都是连续的，这意味着导数过程的功率一定是有限的：

$$R_{X'X'}(t,t) = \mathcal{E}[(X'(t))^2] < \infty \tag{7-129}$$

对一个广义平稳过程，令 $\tau \triangleq t_2 - t_1$（见习题 7-19）：

$$R_{X'X}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau} R_{XX}(\tau), \quad R_{XX'}(\tau) = -\frac{\mathrm{d}}{\mathrm{d}\tau} R_{XX}(\tau) \tag{7-130}$$

$$R_{X'X'}(\tau) = -\frac{\mathrm{d}}{\mathrm{d}\tau} R_{X'X}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau} R_{XX'}(\tau) = -\frac{\mathrm{d}^2}{\mathrm{d}\tau^2} R_{XX}(\tau) \tag{7-131}$$

当自相关函数在 $t_1 = t_2 = t$（或对广义平稳的情况 $\tau = 0$）处不连续，这意味着 MS 导数不存在，狄拉克 δ 函数可以用来代表 $R_{XX}(t_1, t_2)$ 的导数。

例 7-20　对随机电报信号：

$$\frac{\mathrm{d}}{\mathrm{d}\tau} R_{XX}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau} \exp(-2\lambda|\tau|) = -2\lambda \exp(-2\lambda|\tau|)\mathrm{sgn}(\tau) \tag{7-132}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\tau^2} R_{XX}(\tau) = 4\lambda^2 \exp(-2\lambda|\tau|) - 4\lambda \exp(-2\lambda|\tau|)\delta(\tau) \tag{7-133}$$

其中用到了 $\mathrm{sgn}^2(\tau) = 1$，符号函数的导数是 $2\delta(\tau)$。随机电报信号没有 MS 导数，图 7-12非常清晰地说明了这一点。 ◀

定义（均方积分）　随机过程的 MS 积分，记为 $Y(t) \triangleq \int_{t_0}^{t} X(t)\mathrm{d}t$ 对 $[t_0, t] \in \mathcal{T}$，定义如下：

$$\lim_{N\to\infty} \mathcal{E}\left[\left(\int_{t_0}^{t} X(t)\mathrm{d}t - \sum_{n=0}^{N-1} X(t_0 + n\Delta_N)\Delta_N\right)^2\right] = 0 \tag{7-134}$$

其中 $\Delta_N \triangleq (t - t_0)/N$ 是积分的尺寸，它将 $[t_0, t]$ 进行了分区。

定义中的求和是一种阶梯近似，就如对非随机函数（见附录 D）的黎曼求和中所描述的一样。积分可用全部的 N 个非重叠的矩形面积来近似，随着 N 的增长，结果会更精确：分区变得越小，矩形越接近函数。观察到 $Y(t)$ 在 t 时刻是一个随机变量，可用它的均值和自相关函数来描述，就如在 MS 导数中用到的。

为了标记方便，定义式(7-134)中求和项如下：

$$Y_N \triangleq \sum_{n=0}^{N-1} X(t_0 + n\Delta_N)\Delta_N \tag{7-135}$$

则式(7-134)变为：

$$\lim_{N\to\infty} \mathcal{E}[(Y(t) - Y_N)^2] = 0 \tag{7-136}$$

由于 MS 积分 $Y(t)$ 可能是未知的，柯西准则可用来确定 MS 积分是否存在：

$$\lim_{M,N\to\infty} \mathcal{E}[(Y_M - Y_N)^2] = 0 \tag{7-137}$$

它可以写为 $X(t)$ 的自相关函数的形式。

定理 7-15　自相关函数为 $R_{XX}(t_1, t_2)$ 的随机过程 $X(t)$ 在 $[t_0, t] \in \mathcal{T}$ 上有 MS 积分存在，当且仅当 $\int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_1, t_2)\mathrm{d}t_1\mathrm{d}t_2$ 存在。

证明：由柯西准则得到式(7-137)的交叉项为：

$$\mathcal{E}[Y_M Y_N] = \mathcal{E}\left[\sum_{m=1}^{M} X(t_m)\Delta t_m \sum_{n=1}^{N} X(t_n)\Delta t_n\right] = \sum_{m=1}^{M}\sum_{n=1}^{N} R_{XX}(t_m, t_n)\Delta t_m \Delta t_n \tag{7-138}$$

等式的右边是一个对确定函数 $R_{XX}(t_1, t_2)$ 的积分的阶梯近似，则取极限得到：

$$\lim_{M,N\to\infty} \mathcal{E}[Y_M Y_N] = \int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2 \tag{7-139}$$

考虑乘积中的一个外项：

$$\begin{aligned}\lim_{M\to\infty} \mathcal{E}[Y_M^2] &= \lim_{M\to\infty} \mathcal{E}\Big[\sum_{m=1}^{M} X(t_m)\Delta t_m \sum_{n=1}^{M} X(t_n)\Delta t_n\Big] \\ &= \lim_{M\to\infty} \sum_{m=0}^{M}\sum_{n=0}^{M} R_{XX}(t_m,t_n)\Delta t_m \Delta t_n \\ &= \int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2\end{aligned} \tag{7-140}$$

同样对$\lim\limits_{M\to\infty}\mathcal{E}[Y_N^2]$有类似的结果。根据式(7-137)将以上结果联合得到零，证明完毕。

由于式(7-140)是一个非随机函数的双重积分，当$[t_0,t]$有限时，MS 积分总是存在。

定理 7-16　如果 $X(t)$是一个广义平稳过程，则 $R_{XX}(t_1,\ t_2)=R_{XX}(\tau)$，在$[t_0,\ t]$上 $X(t)$的 MS 积分存在，当且仅当下式存在：

$$2\int_{0}^{t-t_0}(t-t_0-\tau)R_{XX}(\tau)\mathrm{d}\tau \tag{7-141}$$

证明：由于广义平稳过程的自相关函数只依赖于唯一一个变量(时间间隔)，我们引入以下的变量替换，设 $\tau\triangleq t_2-t_1$，$\alpha\triangleq t_1+t_2$，得到以下的雅克比变换(见附录 G)：

$$J=\det(\boldsymbol{J})=\det\begin{pmatrix}\partial\tau/\partial t_1 & \partial\tau/\partial t_2\\ \partial\alpha/\partial t_1 & \partial\alpha/\partial t_2\end{pmatrix}=\det\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}=-2 \tag{7-142}$$

则 $\mathrm{d}t_1\mathrm{d}t_2=(1/|J|)\mathrm{d}\tau\mathrm{d}\alpha=(1/2)\mathrm{d}\tau\mathrm{d}\alpha$，因此式(7-140)变为：

$$\begin{aligned}\int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2 &= (1/2)\int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_2-t_1)\mathrm{d}t_1\mathrm{d}t_2 \\ &= \int_{-(t-t_0)}^{t-t_0}\int_{-(t-t_0-|\tau|)}^{t-t_0-|\tau|} R_{XX}(\tau)\mathrm{d}\alpha\mathrm{d}\tau\end{aligned} \tag{7-143}$$

由于是对一个常数内部积分，表达式可简化为：

$$\begin{aligned}\int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2 &= \int_{-(t-t_0)}^{t-t_0}(t-t_0-|\tau|)R_{XX}(\tau)\mathrm{d}\tau \\ &= 2\int_{0}^{t-t_0}(t-t_0-\tau)R_{XX}(\tau)\mathrm{d}\tau\end{aligned} \tag{7-144}$$

其中最后一个表达式用到了 $R_{XX}(\tau)$是偶函数的特性。

当 $X(t)$为广义平稳时，$R_{XX}(t_1,t_2)$对所有$\{t_1,t_2\}$呈现相同的值，可标记为 $\tau=t_2-t_1$。如图 7-16 所示，$R_{XX}(t_1,t_2)$在沿着相对 t_1-t_2 轴逆时针旋转 45°的直线上具有相同值。图中 $R_{XX}(\tau)$相当于任意“薄片”穿过 $R_{XX}(t_1,t_2)$，垂直于那些具有常数相关性的直线，如图所示。观察到在求导中$(t-t_0-|\tau|)$是$[-(t-t_0),t-t_0]$上一个三角函数。初始的积分在$[t_0,t]\times[t_0,t]$定义的正方形中进行。由于 $R_{XX}(t_1,t_2)$沿着垂直于 $t_1=t_2$ 的直线是常数，积分可代之以在 $\tau\in[-(t-t_0),t-t_0]$上进行。但是，为了得到相同的结果，对每个 τ，$R_{XX}(\tau)$需要用正方形中旋转直线的长度给出的值来进行加权。如图中所见，这些线构成了一个三角形，最大值在 $\tau=0$ 处，在 $\tau\in[-(t-t_0),t-t_0]$外则为零。三角加权只是一个对变量变化的“加工”，也就是说，除了 $R_{XX}(\tau)$只依赖于一个变量外，不依赖于随机过程 $X(t)$的任意其他固有特性。

最后，注意到 MS 积分是一个随机变量，具有以下均值：

$$\mathcal{E}\Big[\int_{t_0}^{t} X(t)\mathrm{d}t\Big]=\int_{t_0}^{t}\mathcal{E}[X(t)]\mathrm{d}t=\int_{t_0}^{t}\mu_X(t)\mathrm{d}t \tag{7-145}$$

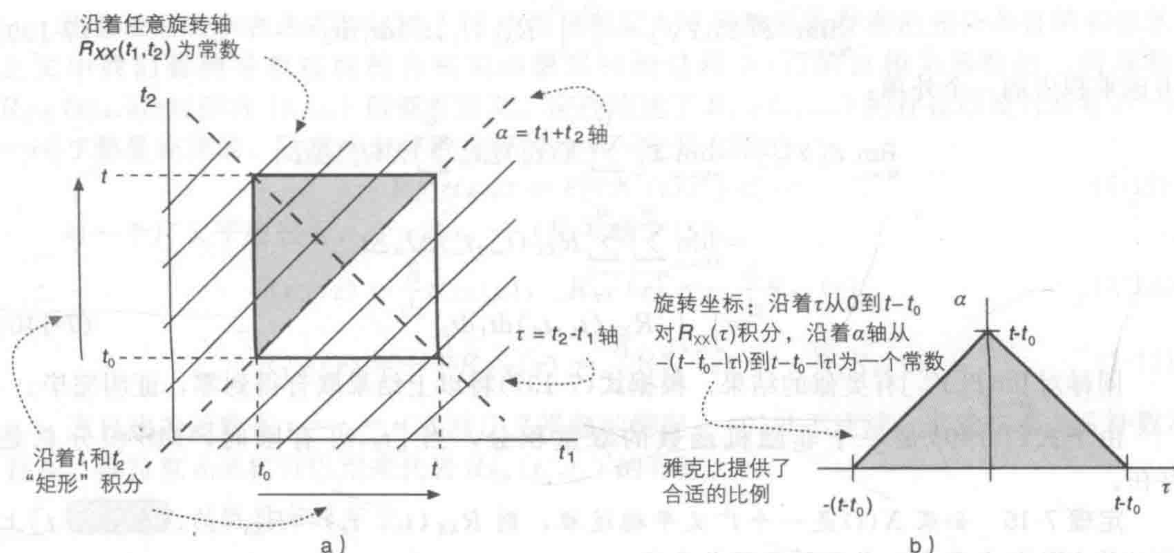


图 7-16 $R_{XX}(t_1, t_2)$的积分范围。a) 初始轴显示了 t_1-t_2 平面上的直线，其中 $R_{XX}(t_1, t_2)=R_{XX}(\tau)$是常数；b) 旋转轴，阴影三角范围显示了积分限

二阶矩：

$$\mathcal{E}\left[\int_{t_0}^{t}\int_{t_0}^{t} X(t_1)X(t_2)\mathrm{d}t_1\mathrm{d}t_2\right]=\int_{t_0}^{t}\int_{t_0}^{t}\mathcal{E}[X(t_1)X(t_2)]\mathrm{d}t_1\mathrm{d}t_2$$

$$=\int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2 \quad (7\text{-}146)$$

方差：

$$\operatorname{var}\left[\int_{t_0}^{t} X(t)\mathrm{d}t\right]=\mathcal{E}\left[\int_{t_0}^{t}\int_{t_0}^{t}(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2))\mathrm{d}t_1\mathrm{d}t_2\right]$$

$$=\int_{t_0}^{t}\int_{t_0}^{t} C_{XX}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2 \quad (7\text{-}147)$$

其中 $C_{XX}(t_1, t_2)$是 $X(t)$的自协方差函数。当 $X(t)$为广义平稳，这些方程简化为：

$$\mathcal{E}\left[\int_{t_0}^{t} X(t)\mathrm{d}t\right]=(t-t_0)\mu_X \quad (7\text{-}148)$$

$$\mathcal{E}\left[\int_{t_0}^{t}\int_{t_0}^{t} X(t_1)X(t_2)\mathrm{d}t_1\mathrm{d}t_2\right]=2\int_{0}^{t-t_0}(t-t_0-\tau)R_{XX}(\tau)\mathrm{d}\tau \quad (7\text{-}149)$$

$$\operatorname{var}\left[\int_{t_0}^{t} X(t)\mathrm{d}t\right]=2\int_{0}^{t-t_0}(t-t_0-\tau)C_{XX}(\tau)\mathrm{d}\tau \quad (7\text{-}150)$$

它们是间隔 $t-t_0$ 的函数。积分过程的二阶矩与用来证明 MS 积分存在的表达式相同，也就是，式(7-146)一定存在。

例 7-21 对随机化电报信号，$C_{XX}(\tau)=\exp(-2\lambda|\tau|)$，因此上式中方差为：

$$\operatorname{var}\left[\int_{t_0}^{t} X(\tau)\mathrm{d}\tau\right]=2\int_{0}^{t-t_0}(t-t_0-\tau)\exp(-2\lambda|\tau|)\mathrm{d}\tau$$

$$=2\exp(-2\lambda(t-t_0))\int_{0}^{t-t_0} v\exp(2\lambda v)\mathrm{d}v \quad (7\text{-}151)$$

我们去掉了指数变量中的绝对值，将变量变为了 $v\triangleq t-t_0-\tau$。积分用分部积分法可得：

$$\int_{0}^{t-t_0} v\exp(2\lambda v)\mathrm{d}v=\left[(v/2\lambda)\exp(2\lambda v)-(1/2\lambda)^2\exp(2\lambda v)\right]\Big|_0^{t-t_0}$$

$$=\exp(2\lambda(t-t_0))[(t-t_0)/2\lambda-1/4\lambda^2]+1/4\lambda^2 \quad (7\text{-}152)$$

将这个结果代入式(7-151)得到：

$$\operatorname{var}\left[\int_{t_0}^{t} X(\tau)\mathrm{d}\tau\right]=(t-t_0)/\lambda+(1/2\lambda^2)[\exp(-2\lambda(t-t_0))-1] \tag{7-153}$$

这也是二阶矩，因为对零均值的随机过程有 $R_{XX}(\tau)=C_{XX}(\tau)$。

总结随机微积分的结果，MS 连续性的定义和随机过程 $X(t)$ 在 $t\in\mathcal{T}$ 的导数需要自相关函数 $R_{XX}(t_1,t_2)$ 在 $t_1=t_2=t$ 处是连续的和可微的，可从 t_1-t_2 平面的任意方向趋近。同样，随机过程 $X(t)$ 在 $[t_0,t]$ 上的 MS 积分需要 $R_{XX}(t_1,t_2)$ 在 $[t_0,t]\times[t_0,t]$ 上的双重积分存在。就如在附录 D 中提到的，关于积分、微分函数的背景，都是连续函数的一个子集，连续函数是可积函数的一个子集。这些关系也用在了之前的 MS 定义中，因为 $R_{XX}(t_1,t_2)$ 是一个非随机函数。下面的定理将详述这些结论。◀

定理 7-17 对 $\mathcal{E}[X^2(t)]<\infty$ 的随机过程 $X(t)$：（ⅰ）如果在 $t\in\mathcal{T}$ 时 MS 导数存在，则 $X(t)$ 在 t 时刻是 MS 连续的；（ⅱ）如果 $X(t)$ 在 $[t_0,t]\in\mathcal{T}$ 上是 MS 连续的，则它在此区间上存在 MS 积分。

证明：（ⅰ）重写 MS 连续的定义中用到的表达式：

$$\mathcal{E}[(X(t+\varepsilon)-X(t))^2]=\varepsilon^2\,\mathcal{E}\left[\left(\frac{X(t+\varepsilon)-X(t)}{\varepsilon}\right)^2\right] \tag{7-154}$$

由于 $X(t)$ 在 t 时刻有 MS 导数，上式右边的期望值的极限存在（是有限的）。因为 $\lim\limits_{\varepsilon\to 0}\varepsilon^2=0$，得到：

$$\lim_{\varepsilon\to 0}\varepsilon^2\lim_{\varepsilon\to 0}\mathcal{E}\left[\left(\frac{X(t+\varepsilon)-X(t)}{\varepsilon}\right)^2\right]=\lim_{\varepsilon\to 0}\varepsilon^2\,\mathcal{E}\left[\left(\frac{X(t+\varepsilon)-X(t)}{\varepsilon}\right)^2\right]=0 \tag{7-155}$$

说明 $X(t)$ 是 MS 连续的。（ⅱ）由习题 7-18 的结果，$R_{XX}(t_1,t_2)$ 对所有 $t_0\leqslant t_1$，$t_2\leqslant t$ 都是连续的，因此 $X(t)$ 的 MS 积分存在。

7.7 微分方程

工程中许多系统都用常系数线性 DE 建模，以描述一些变量（信号、电压、电流等）是如何随时间改变的。在下面讨论中，为了方便起见，我们给出一个简单的二阶 RLC 电路的例子。

例 7-22 一个并联 RLC 电路（电阻 R，电感 L 和电容 C）有独立的电流源 i_s，在 $t=0$ 时刻接入电路。电感上的电流可用以下二阶 DE 来建模：

$$a_2\frac{\mathrm{d}^2}{\mathrm{d}t^2}i(t)+a_1\frac{\mathrm{d}}{\mathrm{d}t}i(t)+a_0 i(t)=i_s u(t) \tag{7-156}$$

其中系数 $\{a_0,a_1,a_2\}$ 是 $\{R,L,C\}$ 的函数，$u(t)$ 是单位阶跃函数。图 7-17 给出了电路图，从中可以得到：

$$a_2=LC,a_1=L/R,a_0=1 \tag{7-157}$$

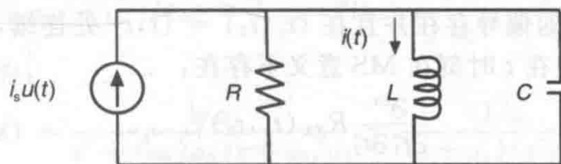


图 7-17 并联 RLC 电路。电流源 i_s，它在 $t=0$ 时刻接入电路，用单位阶跃函数 $u(t)$ 表示 ◀

根据系数值，式(7-156)的解可为以下三种形式之一：

过阻尼： $i(t)=[c_1\exp(-\alpha_1 t)+c_2\exp(-\alpha_2 t)+d]u(t)$ (7-158)

欠阻尼： $i(t)=[(c_1\cos(\omega_d t)+c_2\sin(\omega_d t))\exp(-\alpha t)+d]u(t)$ (7-159)

临界阻尼： $$i(t)=[c_1\exp(-\alpha_1 t)+c_2 t\exp(-\alpha_2 t)+d]u(t) \tag{7-160}$$

其中 $\{\alpha,\alpha_1,\alpha_2\}$ 和 ω_d 是 $\{a_0,a_1,a_2\}$ 的函数，$\{c_1,c_2\}$ 是常数，由电路的初始条件 $\{i(0),i'(0)\}$ 确定，d 是稳态电流（当 $t\to\infty$）。注意到 $\{c_1,c_2\}$ 的值对这三种情况是不同的。通常对这样一个线性系统，叠加定理可用来将以上任意表达式重写为如下形式：

$$i(t)=i_{\mathrm{NR}}(t)+i_{\mathrm{FR}}(t) \tag{7-161}$$

其中 $i_{\mathrm{NR}}(t)$ 是自然响应(NR)，$i_{\mathrm{FR}}(t)$ 是强迫响应(FR)，它们也分别称为零输入响应和零状态响应。将式(7-156)的右边置零得到自然响应，从而得到齐次 DE。强迫响应忽略因储能元件 {L,C} 导致的初始条件，它是基于下文描述的系统的冲激响应。由于这个并联电路有一个电流源，并且电感最终相当于短接，我们得到所有三种情况中有 $d=i_s$：这是当指数项衰减为零时的稳态电流。图 7-18 中显示了三种类型的响应。

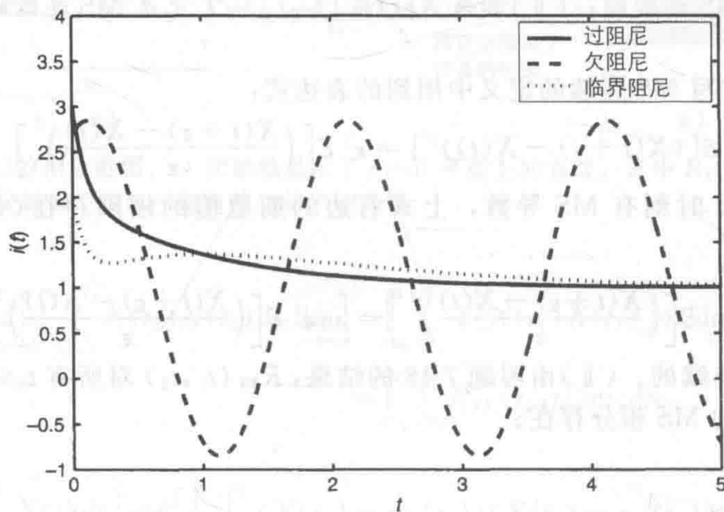


图 7-18 并联 RLC 电路响应的三种类型的举例

我们想将这种对一个线性系统的描述扩展运用到针对随机过程的系统。考虑以下的 n 阶线性随机 DE：

$$a_n\frac{\mathrm{d}^n}{\mathrm{d}t^n}X(t)+a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}X(t)+\cdots+a_1\frac{\mathrm{d}}{\mathrm{d}t}X(t)+a_0X(t)=Y(t) \tag{7-162}$$

其中 $\{a_k\}$ 是常系数。由于 $X(t)$ 和 $Y(t)$ 是随机过程，我们不知道如何求解这样一个系统，因为 $Y(t)$ 和 $X(t)$ 是现实的总体。但是，基于以前 MS 导数的定义，假设对每个 $t\in\boldsymbol{T}$，上式在 MS 意义下成立，我们可以确定均值和相关函数是如何随时间变化的。通常，n 阶 DE 需要 n 个初始条件：这些由随机变量 $\{X(t_0),X^{(1)}(t_0),\cdots,X^{(n-1)}(t_0)\}$ 提供，上标表示微分的阶数。如果以下的偏导存在并且在 $\{t_1,t_2\}=\{t,t\}$ 处连续，扩展 MS 导数，n 阶导数 $\mathrm{d}^nX(t)/\mathrm{d}t^n\triangleq X^{(n)}(t)$ 在 t 时刻在 MS 意义下存在，

$$\frac{\partial^{2n}}{\partial t_1^n\partial t_2^n}R_{XX}(t_1,t_2)\big|_{t_1=t_2=t} \tag{7-163}$$

由于 DE 是常系数线性系统，$Y(t)$ 可看做是一个线性系统的输入，输出为 $X(t)$。这样一个系统可表示如下：

$$X(t)=X_{\mathrm{NR}}(t)+\int_{t_0}^{t}h(t-\tau)Y(\tau)\mathrm{d}\tau \tag{7-164}$$

其中 $h(t)$ 是非随机冲激响应，$X_{\mathrm{NR}}(t)$ 是由非零初始条件引起的自然响应。等式右边的第二项是强迫响应，我们记为 $X_{\mathrm{FR}}(t)$。卷积积分假设在 MS 意义下存在。

我们强调系统本身是非随机的：假设系数 $\{a_m\}$ 是常数。则由于系统是线性时不变的(LTI)，如果输入发生时移，冲激响应不会改变。自然响应 $X_{\mathrm{NR}}(t)$ 可用初始条件 $\{X(t_0),$

$X^{(1)}(t_0),\cdots,X^{(n-1)}(t_0)\}$ 表示如下：

$$X_{NR}(t)=\sum_{m=0}^{n-1}g_m(t)X^{(m)}(t_0)u(t) \tag{7-165}$$

其中 $g_m(t)$是当 $X^{(m)}(t_0)=1$ 其他的初始条件为零时系统的确定响应。因为系统是线性的，所以这种分解是可行的，可以应用叠加定理。在第8章，我们将这些结果扩展到频域，用功率谱密度来描述一个随机过程经一个 LTI 系统滤波后输出 $X(t)$的自相关函数与输入 $Y(t)$的自相关函数的关系。

例 7-23 之前的例子中电路在确定信号输入时，对齐次 DE 的过阻尼类型，$t_0=0$ 的初始条件与系数的相互关系如下：

$$i(0)=c_1+c_2,\quad i'(0)=-\alpha_1c_1-\alpha_2c_2 \tag{7-166}$$

因此：

$$c_1=[\alpha_2i(0)+i'(0)]/(\alpha_2-\alpha_1) \tag{7-167}$$

$$c_2=[-\alpha_1i(0)-i'(0)]/(\alpha_2-\alpha_1) \tag{7-168}$$

其中对齐次 DE 的解为 $d=0$。(理论上，初始条件在 $t=0^+$ 时刻定义，这样就包括了$t=0$时刻的任意不连续，比如用单位阶跃函数建模的电流源。)观察到如果 $i(0)=1$，$i'(0)=0$，则 $c_1=\alpha_2/(\alpha_2-\alpha_1)$，$c_2=-\alpha_1/(\alpha_2-\alpha_1)$；同样，如果 $i(0)=0$，$i'(0)=1$，则 $c_1=1/(\alpha_2-\alpha_1)$，$c_2=-1/(\alpha_2-\alpha_1)$。由这些结果，可得到函数：

$$g_0(t)=[\alpha_2\exp(-\alpha_1t)-\alpha_1\exp(-\alpha_2t)]/(\alpha_2-\alpha_1) \tag{7-169}$$

$$g_1(t)=[\exp(-\alpha_1t)-\exp(-\alpha_2t)]/(\alpha_2-\alpha_1) \tag{7-170}$$

分别用 $i(0)$和 $i'(0)$加权，在式(7-165)中给出了自然响应：

$$\begin{aligned}i_{NR}(t)=&[1/(\alpha_2-\alpha_1)]([i(0)\alpha_2+i'(0)]\exp(-\alpha_1t)\\&-[i(0)\alpha_1+i'(0)]\exp(-\alpha_2t))u(t)\end{aligned} \tag{7-171}$$ ◀

为了完善自然响应的表达式，我们需要这个特殊电路的 $i(0)$和 $i'(0)$。电容和电感的伏安关系如下：

$$i_C(t)=C\frac{\mathrm{d}v_C(t)}{\mathrm{d}t},\quad v_L(t)=L\frac{\mathrm{d}i_L(t)}{\mathrm{d}t} \tag{7-172}$$

这意味着 $v_C(t)$和 $i_L(t)$不会发生跃变。如果对于 $t<0$，所有的电压、电流都为零，则 $i_L(0)=0$，$v_C(0)=0$，由 $i(0)=0$ 和 $i'(0)=0\Rightarrow v_L(0)=0$。因为在这个例题中初始条件为零，自然响应 $i_{NR}(t)=0$。

通常做系统分析时，初始条件一般都假设为零，只需考虑用系统的冲激响应 $h(t)$来表征的强迫响应。传递函数 $H(s)$由 DE 的拉普拉斯变化推导得到：

$$(a_2s^2+a_1s+a_0)I(s)=I_s/s\Rightarrow H(s)=\frac{1}{a_2s^2+a_1s+a_0} \tag{7-173}$$

将这个表达式重写为：

$$H(s)=\frac{1/a_2}{s^2+(a_1/a_2)s+a_0/a_2}=\frac{1/a_2}{(s+\alpha_1)(s+\alpha_2)} \tag{7-174}$$

其中 $\{\alpha_1,\alpha_2\}$ 为 $H(s)$的极点。由于本例中 $a_0=1$，$a_2=1/\alpha_1\alpha_2$ 以及 $a_1=(\alpha_1+\alpha_2)/\alpha_1\alpha_2$。

由部分分数展开(PFE)，得到拉普拉斯反变换：

$$h(t)=\frac{\alpha_1\alpha_2}{\alpha_2-\alpha_1}[\exp(-\alpha_1t)-\exp(-\alpha_2t)]u(t) \tag{7-175}$$

由于输入为一个阶跃函数，强迫响应可简化为：

$$i_{FR}(t)=I_s\int_0^t h(t-\tau)\mathrm{d}\tau \tag{7-176}$$

代入冲激响应得到：

$$\begin{aligned} i_{\mathrm{FR}}(t) &= \frac{\alpha_1\alpha_2 I_s}{\alpha_2-\alpha_1}\left[\int_0^t \exp(-\alpha_1(t-\tau))\mathrm{d}\tau - \int_0^t \exp(-\alpha_2(t-\tau))\mathrm{d}\tau\right]u(t) \\ &= \frac{\alpha_2 I_s}{\alpha_2-\alpha_1}[1-\exp(-\alpha_1 t)]u(t) - \frac{\alpha_1 I_s}{\alpha_2-\alpha_1}[1-\exp(-\alpha_2 t)]u(t) \\ &= \left[\frac{\alpha_1 I_s}{\alpha_2-\alpha_1}\exp(-\alpha_2 t) - \frac{\alpha_2 I_s}{\alpha_2-\alpha_1}\exp(-\alpha_1 t) + I_s\right]u(t) \end{aligned} \tag{7-177}$$

由于 $i_{\mathrm{NR}}(t)=0$，所以这也是全响应。这个表达式与式(7-158)具有相同的形式，因此：

$$c_1 = -\frac{\alpha_2 I_s}{\alpha_2-\alpha_1}, c_2 = \frac{\alpha_1 I_s}{\alpha_2-\alpha_1}, d = i_s \tag{7-178}$$

显然正如预期，极限 $\lim_{t\to 0} i_{\mathrm{FR}}(t) = i_s$。

如果例 7-23 中输入是一个随机过程，则会得到同样的 $\{g_m(t)\}$，因为它们是确定的。但是，自然响应是随机的，因为式(7-165)中的初始条件是随机变量。正如之前提到的，线性 DE 的导数和卷积积分都是在 MS 意义下定义的。

接下来，我们给出用 DE 来描述的随机过程 $X(t)$ 的均值和自相关函数的特征。我们将会发现得到的方程与之前电路中确定信号的结果有些类似。式(7-162)中 DE 的 m 阶均值函数为：

$$\mu_X^{(m)}(t) \triangleq \mathcal{E}\left[\frac{\mathrm{d}^m}{\mathrm{d}t^m}X(t)\right] = \frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{E}[X(t)] \tag{7-179}$$

将这个结果应用到整个 DE 得到：

$$\sum_{m=0}^{n} a_m \mu_X^{(m)}(t) = \mu_Y(t) \tag{7-180}$$

为了解这个方程，需要以下的初始条件 $\{\mu_X^{(n-1)}(0),\cdots,\mu_X^{(1)}(0),\mu_X(0)\}$。交换求导和求期望的顺序，得到以下的互相关函数：

$$\begin{aligned} R_{XY}(t_1,t_2) &= \mathcal{E}[X(t_1)Y(t_2)] \\ &= \mathcal{E}\left[X(t_1)\sum_{m=0}^{n} a_m \frac{\partial^m}{\partial t_2^m}X(t_2)\right] = \sum_{m=0}^{n} a_m \frac{\partial}{\partial t_2}R_{XX}(t_1,t_2) \end{aligned} \tag{7-181}$$

其中 t_1，$t_2 \geqslant 0$。这是另一个线性 DE，但是以互相关函数而不是随机过程来表征。与式(7-162)比较，我们看到对任意固定的 t_1，它可看为一个参数，上式中的 DE 是一个 t_2 的函数。同样可以写为：

$$R_{YY}(t_1,t_2) = \sum_{m=0}^{n} a_m \frac{\partial^m}{\partial t_1^m} R_{XY}(t_1,t_2) \tag{7-182}$$

它是一个 $t_1 \geqslant 0$ 的函数，以 $t_2 \geqslant 0$ 为参量。式(7-182)中 $R_{YY}(t_1,t_2)$ 的答案可以代入式(7-181)中，这样给定初始条件就可能解出 $R_{XX}(t_1,t_2)$。第 8 章中，我们将证明用傅里叶和拉普拉斯变换求解输入和输出的自相关函数更简单。例 7-24 给出了一个示例。

例 7-24 考虑以下二阶 DE：

$$a_2\frac{\mathrm{d}^2}{\mathrm{d}t^2}X(t) + a_1\frac{\mathrm{d}}{\mathrm{d}t}X(t) + a_0 X(t) = Y(t) \tag{7-183}$$

对于 $t>0$，$Y(t)$ 是一个零均值过程，自相关函数为 $R_{YY}(\tau)=\delta(\tau)$。假设 $\{a_0,a_1,a_2\}$ 已选定，即系统为过阻尼，由式(7-179)观察得到：

$$a_2\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mu_X(t) + a_1\frac{\mathrm{d}}{\mathrm{d}t}\mu_X(t) + a_0\mu_X(t) = \mu_Y(t) = 0 \tag{7-184}$$

是一个非受迫 DE。由式(7-158)($d=0$)，答案具有以下形式：

$$\mu_X(t) = [c_1\exp(-\alpha_1 t) + c_2\exp(-\alpha_2 t)]u(t) \tag{7-185}$$

其中 $\{c_1,c_2\}$ 由初始条件 $\{\mu'_X(0),\mu_X(0)\}$ 确定，$\{\alpha_1,\alpha_2\}$ 是特征方程的实数根：

$$a_2 s^2 + a_1 s + a_0 = 0 \tag{7-186}$$

在式(7-184)中应用单边拉普拉斯变换，代入初始条件得到：

$$a_2 s^2 \mu_X(s) - a_2 s\mu_X(0) - a_2\mu'_X(0) + a_1 s\mu_X(s) - a_1\mu_X(0) + a_0\mu_X(s) = 0 \tag{7-187}$$

由部分分数展开(PFE)得到：

$$\mu_X(s) = [a_2\mu'_X(0) + (a_2 s + a_1)\mu_X(0)]/(a_2 s^2 + a_1 s + a_0) = \frac{c_1}{s+\alpha_1} + \frac{c_2}{s+\alpha_2} \tag{7-188}$$

由拉普拉斯反变化得到：

$$\mu_X(t) = c_1\exp(-\alpha_1 t)u(t) + c_2\exp(-\alpha_2 t)u(t) \tag{7-189}$$

当然，如果初始条件为零，则对 $t \geqslant 0$，有 $\mu_X(t)=0$。对非零初始条件，假设 α_1，$\alpha_2 >$ 0，则 $\mu_X(t) \to 0$。因此，即使输入 $Y(t)$ 广义平稳，输出 $X(t)$ 也不是广义平稳的，因为均值为时变的。但是，由于均值是收敛的，如果 $R_{XX}(t_1,t_2) \to R_{XX}(\tau)$，则 $X(t)$ 是渐进广义平稳的。这种非平稳是因为电流源 I_s 在 $t=0$ 时刻开关，而且在 {L,C} 中有能量存储(对应于非零初始条件)。 ◀

对互相关函数，由式(7-182)得到：

$$a_2\frac{\partial^2}{\partial t_1^2}R_{XY}(t_1,t_2) + a_1\frac{\partial}{\partial t_1}R_{XY}(t_1,t_2) + a_0 R_{XY}(t_1,t_2) = R_{YY}(t_1,t_2) = \delta(t_1 - t_2) \tag{7-190}$$

其中 t_1，$t_2 > 0$。正如之前提到的，这个 DE 的解答是一个 t_1 的函数，t_2 被看做一个参数。由于等式的右边是狄拉克 δ 函数，解答是 DE 代表的系统的冲激响应函数 $h(t_1 - t_2)$。它是一个 t_1 的函数，向右移动了 t_2，因为 δ 函数有同样的偏移。应用拉普拉斯变换(只对 t_1 操作)得到：

$$\begin{aligned} &a_2 s^2 R_{XY}(s,t_2) - a_2 sR_{XY}(0,t_2) - a_2 R'_{XY}(0,t_2) + a_1 R_{XY}(s,t_2) \\ &\quad - a_1 R_{XY}(0,t_2) + a_0 R_{XY}(s,t_2) = \exp(-st_2) \end{aligned} \tag{7-191}$$

其中用到了拉普拉斯变换的偏移特性得到等式右边的结果(见附录C)。由于在式(7-191)中假设 t_2 是固定的，可用以下导数的记号指定 $t_1=0$ 时刻的初始条件：

$$R'_{XY}(0,t_2) \triangleq \frac{\partial}{\partial t_1}R_{XY}(t_1,t_2)\big|_{t_1=0} \tag{7-192}$$

则：

$$R_{XY}(s,t_2) = [\exp(-st_2) + a_2 R'_{XY}(0,t_2) + (a_2 s + a_1)R_{XY}(0,t_2)]/(a_2 s^2 + a_1 s + a_0) \tag{7-193}$$

假设零初始条件 $R'_{XY}(0,t_2) = R_{XY}(0,t_2) = 0$，强迫响应的拉普拉斯变换为：

$$R_{XY}(s,t_2) = \exp(-st_2)\left(\frac{c_3}{s+\alpha_1} + \frac{c_4}{s+\alpha_2}\right) \tag{7-194}$$

其中 $\{c_3, c_4\}$ 与求均值 $\mu_X(t)$ 时得到的不同。拉普拉斯反变换得到：

$$R_{XY}(t_1,t_2) = [c_3\exp(-\alpha_1(t_1 - t_2)) + c_4\exp(-\alpha_2(t_1 - t_2))]u(t_1 - t_2) \tag{7-195}$$

它是一个时间差 $t_1 - t_2$ 的函数。但是，注意到这是一个互相关函数，它并不意味着 $X(t)$ 是广义平稳的；我们需要推导 $R_{XX}(t_1,t_2)$ 的表达式来确定平稳性。对固定的 t_2，$R_{XY}(t_1,t_2)$ 是一个 t_1 的右函数，只有 $t_1 > t_2$ 时非零，并且在 $\mathcal{R}^+$ 上是指数衰减到零，如图 7-19a所示。对固定的 t_1，函数可被看做一个在 $[0,t_1)$ 区间指数增长的函数。当要计算式(7-181)中的 DE 时，通常采用后一种观点，DE 变为：

$$a_2\frac{\partial^2}{\partial t_2^2}R_{XX}(t_1,t_2) + a_1\frac{\partial}{\partial t_2}R_{XX}(t_1,t_2) + a_0 R_{XX}(t_1,t_2) = R_{XY}(t_1,t_2) \tag{7-196}$$

对这个二阶的例子，$R_{XY}(t_1,t_2)$ 是式(7-195)中的解答。由于 t_1 被看做一个参数，这个方程

描述了对每个t_1值$R_{XX}(t_1,t_2)$如何随t_2改变。假设有零初始条件，由拉普拉斯变换可得到以下的强迫响应：

$$R_{XX}(t_1,s)=R_{XY}(t_1,s)/[a_2s^2+a_1s+a_0]=R_{XY}(t_1,s)\left(\frac{c_5}{s+\alpha_1}+\frac{c_6}{s+\alpha_2}\right) \quad (7\text{-}197)$$

其中$\{c_5,c_6\}$与之前的系数都不同，尽管极点$\{\alpha_1,\alpha_2\}$与之前相同(它们给出了DE的特征，与初始条件无关)。式(7-197)右边的输入$R_{XY}(t_1,s)$是$R_{XY}(t_1,t_2)$关于t_2的拉普拉斯变换。如果不做变换，我们将式(7-195)与式(7-197)中的系统的冲激响应卷积，系统冲激响应为：

$$h(t_2)=[c_5\exp(-\alpha_1t_2)+c_6\exp(-\alpha_2t_2)]u(t_2) \quad (7\text{-}198)$$

卷积为：

$$\begin{aligned}R_{XX}(t_1,t_2)&=\int_0^\infty R_{XY}(t_1,\tau)h(t_2-\tau)\mathrm{d}\tau\\&=\int_0^{\min(t_1,t_2)}c_3c_5\exp(-\alpha_1(t_1-\tau)\exp(-\alpha_1(t_2-\tau))\mathrm{d}\tau\\&\quad+\int_0^{\min(t_1,t_2)}c_3c_6\exp(-\alpha_1(t_1-\tau)\exp(-\alpha_2(t_2-\tau))\mathrm{d}\tau\\&\quad+\int_0^{\min(t_1,t_2)}c_4c_5\exp(-\alpha_2(t_1-\tau)\exp(-\alpha_1(t_2-\tau))\mathrm{d}\tau\\&\quad+\int_0^{\min(t_1,t_2)}c_4c_6\exp(-\alpha_2(t_1-\tau)\exp(-\alpha_2(t_2-\tau))\mathrm{d}\tau\end{aligned} \quad (7\text{-}199)$$

其中阶跃函数确定了每个积分的上限。第一项为：

$$\begin{aligned}&c_3c_5\exp(-\alpha_1(t_1+t_2))\int_0^{\min(t_1,t_2)}\exp(2\alpha_1\tau)\mathrm{d}\tau\\&=\frac{c_3c_5}{2\alpha_1}[\exp(-\alpha_1|t_1-t_2|)-\exp(-\alpha_1(t_1+t_2))]\end{aligned} \quad (7\text{-}200)$$

其中用到了$\min(t_1, t_2)=(1/2)(t_1+t_2-|t_1-t_2|)$。可将以上结果中的$c_3c_5$替换为$c_4c_6$，$\alpha_1$替换为$\alpha_2$就可以得到第四项。第二和第三项比较复杂些，但和其他两项具有相同的基本形式：一是为$|t_1-t_2|$的函数，另一个是t_1和t_2的加权和(见习题7-22)。因此，我们发现尽管$X(t)$是非平稳的，但当t_1，$t_2\to\infty$时，它是渐进广义平稳的：

$$\begin{aligned}R_{XX}(t_1,t_2)\to&\frac{c_3c_5}{2\alpha_1}\exp(-\alpha_1|t_1-t_2|)+\frac{c_4c_6}{2\alpha_2}\exp(-\alpha_2|t_1-t_2|)\\&+\frac{c_3c_6+c_4c_5}{\alpha_1+\alpha_2}\exp(-(\alpha_1+\alpha_2)|t_1-t_2|/2)\end{aligned} \quad (7\text{-}201)$$

其中只有$|t_1-t_2|$的指数项被保留了(例如含有t_1+t_2项的都趋于零了)。由图7-19b和c中的例子可以很清晰的看出非平稳性，我们看到当$t_1=t_2=t$对应t_1和t_2在零附近时，$R_{XX}(t_1,t_2)$发生了变化。

例7-24要复杂些，因为相关函数有两个变量$\{t_1,t_2\}$。在第8章，我们将关注广义平稳随机过程，它们的相关函数是单一变量τ的函数。这将简化以上的分析，由DE建模的线性系统的特性就更容易用功率谱密度(PSD)在频域中分析。分析类似于对确定信号的分析，但此处的"信号"是一个随机过程的均值和相关函数。

最后，我们提到式(7-162)中的DE可以扩展为包括输入的导数，如下所示：

$$\sum_{m=0}^{n_a}a_m\frac{\mathrm{d}^m}{\mathrm{d}t^m}X(t)=\sum_{m=0}^{n_b}b_m\frac{\mathrm{d}^m}{\mathrm{d}t^m}Y(t) \quad (7\text{-}202)$$

之前所有的结果都可以扩展运用到这里。就如在DE描述确定信号中所做的，用拉普拉斯变换得到s域中对应的表达式。对式(7-202)中的一般的DE，转移函数是具有常系数的s中的两个多项式的比值：它是s的一个有理函数。

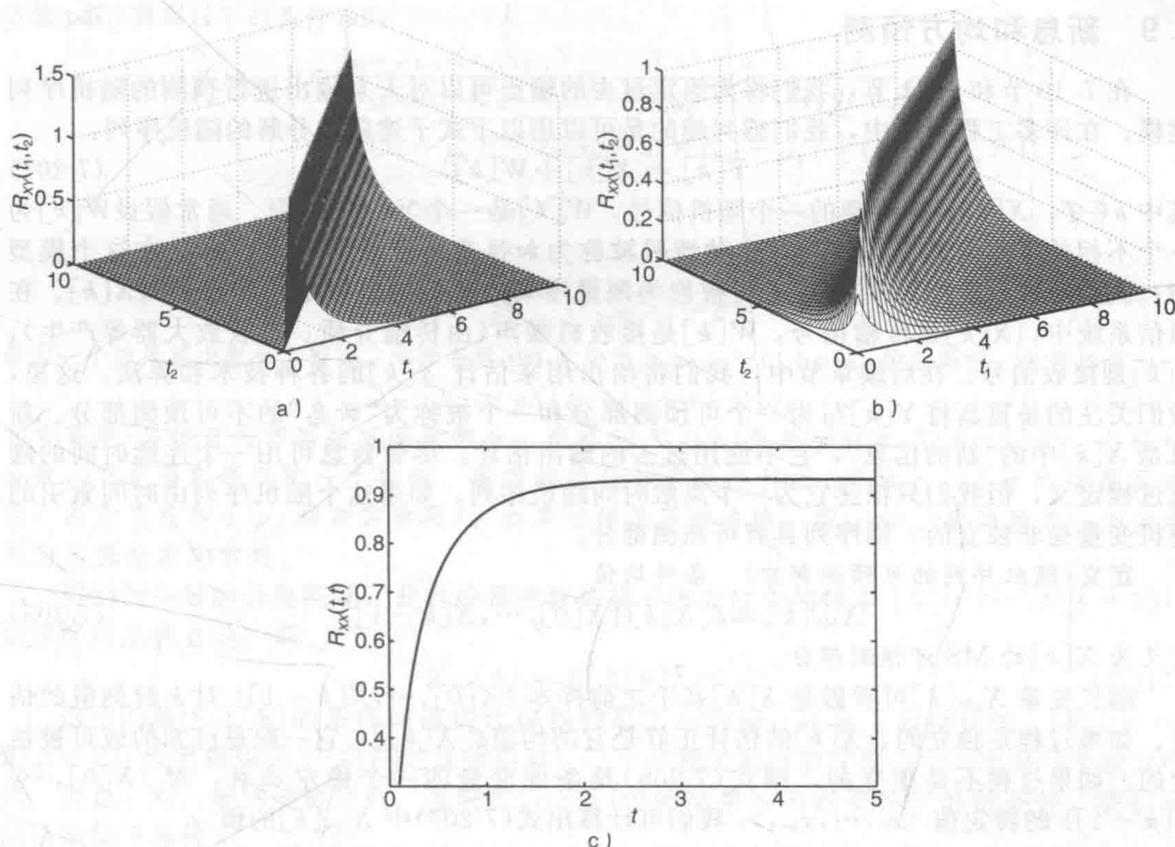


图 7-19 例 7-24 中的相关函数。a）互相关函数 $R_{XY}(t_1,t_2)$；b）自相关函数 $R_{XX}(t_1,t_2)$；c）对应 $t_1=t_2=t$ 的自相关函数 $R_{XX}(t_1,t_2)$

7.8 差分方程

对一个用以下线性差分方程表征的离散时间系统，随机序列 $Y[k]$ 经过这个系统后可推导得到相同的结果：

$$\sum_{m=0}^{n_a} a_m X[k-m] = \sum_{m=0}^{n_b} b_m Y[k-m] \tag{7-203}$$

得到输出 $X[k]$。取期望值得到以下有关均值的方程：

$$\sum_{m=0}^{n_a} a_m \mu_X[k-m] = \sum_{m=0}^{n_b} b_m \mu_Y[k-m] \tag{7-204}$$

也可直接生成各种相关函数的类似的方程：

$$\sum_{m=0}^{n_a} a_m R_{XX}[k_1-m,k_2] = \sum_{m=0}^{n_b} b_m R_{YX}[k_1-m,k_2] \tag{7-205}$$

$$\sum_{m=0}^{n_a} a_m R_{XY}[k_1-m,k_2] = \sum_{m=0}^{n_b} b_m R_{YY}[k_1-m,k_2] \tag{7-206}$$

其中随机序列不一定需要广义平稳。第一个方程通过用式(7-203)的 k_1 代替 k 得到，然后乘以 $X[k_2]$，再取期望值。第二个方程用同样的方法得到，但这里是乘以 $Y[k_2]$。这些差分方程可用之前 DE 例子中用到的类似的方法来分析，但此处是用 z 变换而不是拉普拉斯变换。我们将这个讨论延到第 8 章，在那里我们将关注广义平稳信号，尽管例 7-24中考虑的是一个一阶的例子。

7.9 新息和均方预测

在 7.10 节和 7.11 节，我们将为给定过去的输出可以对未来输出进行预测的随机序列建模。在许多工程问题中，我们感兴趣的是可以用以下式子建模或分解的随机序列：

$$Y[k]=X[k]+W[k] \tag{7-207}$$

其中 $k\in\mathcal{T}$，$X[k]$是感兴趣的一个随机信号，$W[k]$是一个“噪声”过程。通常假设$W[k]$为一个不相关零均值的高斯序列，这种情况被称为加性高斯白噪声(AWGN)。在这个模型中，我们通常只能得到$Y[k]$，有时被称为测量序列，我们的目标是估计或预测$X[k]$。在通信系统中，$X[k]$是传输信号，$W[k]$是接收机噪声(由传输介质、接收放大器等产生)，$Y[k]$是接收信号。在后续章节中，我们将给出用来估计 $X[k]$的各种技术和算法。这里，我们关注的是重新将$Y[k]$写为一个可预测部分和一个被称为“新息”的不可预测部分。新息是$X[k]$中的“新的信息”，它不能用过去的输出估计。尽管新息可用一个连续时间的随机过程定义，但我们只描述它为一个离散时间随机序列。如果这个随机序列由时间索引的随机变量是非独立的，则序列具有可预测部分。

定义(随机序列的可预测部分) 条件均值

$$X_{\mathrm{MS}}[k]\triangleq\mathcal{E}[X[k]\,|\,X[0],\cdots,X[k-1]] \tag{7-208}$$

定义为$X[k]$的 MS 可预测部分。

随机变量 $X_{\mathrm{MS}}[k]$可看做是$X[k]$基于之前样本$\{X[0],\cdots,X[k-1]\}$对k时刻值的估计。如果过程是独立的，$X[k]$的估计正好是它的均值$\mathcal{E}[X[k]]$，它一般是已知的或可被估计的。如果过程不是独立的，则式(7-208)是条件变量的一个确定函数：对$\{X[0],\cdots,X[k-1]\}$的特定值$\{x_0,\cdots,x_{k-1}\}$，我们可计算出式(7-208)中$X_{\mathrm{MS}}[k]$的值。

回顾第 5 章，条件均值是最小 MS 误差(MMSE)估计，对这种情况 MSE 为：

$$\mathcal{E}[(X[k]-g(X[0],\cdots,X[k-1]))^2] \tag{7-209}$$

$g(\cdot)$是上面提到的确定函数，用来由之前所有样本估计当前的样本。对特定的条件值，条件均值计算如下：

$$\begin{aligned}&\mathcal{E}[X[k]\,|\,X[0]=x_0,\cdots,X[k-1]=x_{k-1}]\\&\quad=\int_{-\infty}^{\infty}x_k f_{X[k]\,|\,X[0],\cdots,X[k-1]}(x_k\,|\,x_0,\cdots,x_{k-1})\mathrm{d}x_k=g(x_0,\cdots,x_{k-1})\end{aligned} \tag{7-210}$$

通常需要随机序列k个时刻的联合 pdf。如果序列正好为马尔可夫序列，条件均值简化为$\mathcal{E}[X[k]\,|\,X[0],\cdots,X[k-1]]=\mathcal{E}[X[k]\,|\,X[k-1]]$，而且

$$\begin{aligned}\mathcal{E}[X[k]\,|\,X[k-1]=x_{k-1}]&=\int_{-\infty}^{\infty}x_k f_{X[k]\,|\,X[k-1]}(x_k\,|\,x_{k-1})\mathrm{d}x_k\\&=\frac{1}{f_{X[k-1]}(x_{k-1})}\int_{-\infty}^{\infty}x_k f_{X[k],X[k-1]}(x_k,x_{k-1})\mathrm{d}x_k\end{aligned} \tag{7-211}$$

因此只需要随机序列在$k-1$和k时刻的二维 pdf。这是可直接得到的：对一个马尔可夫过程，只需要最近样本$X[k-1]$的信息，这意味着预测只需要$X[k-1]$和$X[k]$的联合 pdf。如果序列不是马尔可夫的，我们仍可以用式(7-211)，但是它被称为一步预测器，因为它只基于最近样本对当前样本的估计，不像式(7-208)需要之前所有样本。

例 7-25 设$X[k]$是一个高斯序列，参数为常数$\{\mu_X,\sigma_X\}$，假设它是一个马尔可夫序列，则我们只需二维 pdf：

$$\begin{aligned}f_{X[k],X[k-1]}(x_k,x_{k-1})=&\frac{1}{2\pi\sigma_X^2\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2\sigma_X^2(1-\rho^2)}\\&\cdot[(x_k-\mu_X)^2-2\rho(x_k-\mu_X)(x_{k-1}-\mu_X)+(x_{k-1}-\mu_X)^2]\Big)\end{aligned} \tag{7-212}$$

其中$\rho\triangleq\mathcal{E}[(X[k]-\mu_X)(X[k-1]-\mu_X)]/\sigma_X^2$是它的相关系数。将这个联合 pdf 除以$k-1$时刻的

边缘 pdf，得到以下的条件 pdf：

$$f_{X[k]\mid X[k-1]}(x_k \mid x_{k-1}) = \frac{f_{X[k],X[k-1]}(x_k, x_{k-1})}{f_{X[k-1]}(x_{k-1})} = \frac{1}{\sqrt{2\pi}\sigma_X}\exp\Big(-\frac{1}{2\sigma_X^2(1-\rho^2)}$$

$$\cdot [(x_k-\mu_X)^2 - 2\rho(x_k-\mu_X)(x_{k-1}-\mu_X) + (x_{k-1}-\mu_X)^2]$$

$$+\frac{1}{2\sigma_X^2}(x_{k-1}-\mu_X)^2\Big)$$

$$= \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}}\exp\Big(-\frac{1}{2\sigma_X^2(1-\rho^2)}[x_k-(1+\rho)\mu_X-\rho x_{k-1}]^2\Big) \quad (7\text{-}213)$$

最后一个结果可由配平方得到。由于条件 pdf 是均值为 $\rho x_{k-1}+(1+\rho)\mu_X$ 的高斯型，条件均值为：

$$X_{\mathrm{MS}}[k] = \rho X[k-1] + (1+\rho)\mu_X \quad (7\text{-}214)$$

它是高斯-马尔可夫序列的可预测部分。观察到 $X_{\mathrm{MS}}[k]$是最近样本 $X[k-1]$的一个线性函数，它是高斯序列的一个特征：条件均值估计器是线性的(实际上，上面的结果是近似，因为它包括了 μ_X 相加常数项)。如果过程是不相关的，即 $\rho=0$，预测器是一个由序列的均值给定的常数。◀

再回到一般的分解问题，此处的预测器是基于所有过去的样本$\{X[0],\cdots,X[k-1]\}$，如果序列是独立的，则

$$X_{\mathrm{MS}}[k] = \mathcal{E}[X[k]] = \mu_X \quad (7\text{-}215)$$

这个问题中 $X[k]$的条件均值估计得到的是它的均值，正如上面提到的：$\{X[0],\cdots,X[k-1]\}$没有提供关于 $X[k]$的任何信息。一个非相关的高斯序列($\rho=0$)也是一个独立序列。但是，对一般的随机序列，得到式(7-215)中的结果并不需要独立性的条件，我们可以给出如下条件。

定义(均方不可预测性)　随机序列 $X[k]$是 **MS 不可预测**的，如果

$$\mathcal{E}[X[k]\mid X[0],\cdots,X[k-1]] = \mathcal{E}[X[k]] \quad (7\text{-}216)$$

MS 不可预测性是一个比独立性弱的条件，但比不相关强。这点通过检查联合期望$\mathcal{E}[X[0],\cdots,X[k-1],X[k]]$就很容易证明。如果 $X[k]$和$\{X[0],\cdots,X[k-1]\}$是不相关的(或独立的)，则期望值分解为：

$$\mathcal{E}[X(0],\cdots,X[k-1],X[k]] = \mathcal{E}[X[0],\cdots,X[k-1]]\,\mathcal{E}[X[k]] \quad (7\text{-}217)$$

这个结果也可通过式(7-216)得到，如下所示：

$$\mathcal{E}[X[0],\cdots,X[k-1],X[k]] = \mathcal{E}[X[0],\cdots,X[k-1],\mathcal{E}[X[k]\mid X[0],\cdots,X[k-1]]]$$

$$= \mathcal{E}[X[0],\cdots,X[k-1],\mathcal{E}[X[k]]]$$

$$= \mathcal{E}[X[0],\cdots,X[k-1]]\,\mathcal{E}[X[k]] \quad (7\text{-}218)$$

第一个等式永远成立(嵌套期望)，第二个等式源于式(7-216)中的条件。最后一个表达式$\mathcal{E}[X(k)]$被放到求期望算子的外面，因为内部期望值是一个常数：它不依赖于外部期望的任意值。由最后的结果，式(7-216)意味着 $X[k]$和$\{X[0],\cdots,X[k-1]\}$是不相关的；但是不相关并不意味着式(7-216)。独立总是意味着式(7-216)，但是反过来通常不成立。为方便起见，我们将随机序列 $X[k]$的期望值的几个特性总结在表 7-4 中。

表 7-4　随机序列 $X[k]$的期望值的特性

特性	期望值的结果
不相关	$\mathcal{E}[X[k-1]X[k]]=\mathcal{E}[X[k-1]]\mathcal{E}[X[k]]$
正交	$\mathcal{E}[X[k-1]X[k]]=0$
均方不可预测	$\mathcal{E}[X[k]\mid X[k-1]]=\mathcal{E}[X[k]]$
一阶马尔可夫	$\mathcal{E}[X[k]\mid X[k-1],\ X[k-2]]=\mathcal{E}[X[k]\mid X[k-1]]$
鞅	$\mathcal{E}[X[k]\mid X[k-1]]=X[k-1]$

给定 $X_{MS}[k]$是 $X[k]$在条件均值下的可预测部分，这意味着有一个不可预测部分，我们称之为新息。

定义(新息)　随机序列 $X[k]$的**新息** $V[k]$是

$$V[k] \triangleq X[k] - X_{MS}[k] \tag{7-219}$$

它是 $X[k]$中新的信息，由随机序列以前的样本无法实现 MS 可预测。

观察到对于 $k>0$，以之前的样本为条件的 $V[k]$的期望值是零：

$$\begin{aligned}\mathcal{E}[V(k] \mid X[0],\cdots,X[k-1]] &= \mathcal{E}[X[k] \mid X[0],\cdots,X[k-1]] \\ &\quad - \mathcal{E}[X_{MS}[k] \mid X[0],\cdots,X[k-1]] \\ &= X_{MS}[k] - X_{MS}[k] = 0\end{aligned} \tag{7-220}$$

由于 $X_{MS}[k]$已经是以之前的样本为条件了，第二个具有同样条件的期望值再次给出了$X_{MS}[k]$，因为它在第二个条件期望下基本上是一个常数。对 $k>0$，$V[k]$的均值也是零：

$$\begin{aligned}\mathcal{E}[V[k]] &= \mathcal{E}[X[k]] - \mathcal{E}[X_{MS}[k]] \\ &= \mathcal{E}[X[k]] - \mathcal{E}[\mathcal{E}[X[k] \mid X[0],\cdots,X[k-1]]] \\ &= \mathcal{E}[X[k]] - \mathcal{E}[X[k]] = 0\end{aligned} \tag{7-221}$$

其中再次用到了嵌套期望值。因此，由式(7-220)和式(7-221)：

$$\mathcal{E}[V[k] \mid X[0],\cdots,X[k-1]] = \mathcal{E}[V[k]] = 0 \tag{7-222}$$

因此 $V[k]$是 $X[k]$的 MS 不可预测部分。而且，可以证明 $V[k]$本身就是 MS 不可预测的。

定理 7-18　随机序列 $X[k]$的新息 $V[k]$是 MS 不可预测的：

$$\mathcal{E}[V[k] \mid V[0],\cdots,V[k-1]] = \mathcal{E}[V[k]] \tag{7-223}$$

证明：这个结果可通过证明下式得到：

$$\mathcal{E}[V[k] \mid V[0],\cdots,V[k-1]] = \mathcal{E}[V[k] \mid X[0],\cdots,X[k-1]] \tag{7-224}$$

因为我们已经证明了等式右边是$\mathcal{E}[V[k]]$。由新息的定义得到：

$$\begin{aligned}V[k-1] &= X[k-1] - \mathcal{E}[X[k-1] \mid X[0],\cdots,X[k-2]] \\ &\vdots \\ V[1] &= X[1] - \mathcal{E}[X[1] \mid X[0]] \\ V[0] &= X[0]\end{aligned} \tag{7-225}$$

这些表达式是新息序列定义的方式。由于新息可被表示为 $\{X[0],\cdots,X[k-1]\}$ 的一个函数，可以用条件期望的平滑特性(见第 5 章)得到下式：

$$\begin{aligned}\mathcal{E}[V[k] \mid X[0],\cdots,X[k-1]] &= \mathcal{E}[\mathcal{E}[V[k] \mid X[0],\cdots,X[k-1]] \mid V[0],\cdots,V[k-1]] \\ &= \mathcal{E}[V[k] \mid V[0],\cdots,V[k-1]]\end{aligned} \tag{7-226}$$

第一个等式成立是因为它是一个嵌套期望。第二个等式成立是因为 $\{V[0],\cdots,V[k-1]\}$ 是由 $\{X[0],\cdots,X[k-1]\}$ 得到的，这意味着它们具有一个粗糙 σ—域。因此，基于 $\{X[0],\cdots,X[k-1]\}$ 的条件可忽略，证明完毕。

定理 7-19　新息 $V[k]$与 MS 可预测部分 $X_{MS}[k]$是正交的：

$$\mathcal{E}[X_{MS}[k]V[k]] = 0 \tag{7-227}$$

证明：代入 $V[k]$的定义得到：

$$\mathcal{E}[X_{MS}[k](X[k] - X_{MS}[k])] = \mathcal{E}[X_{MS}[k]X[k]] - \mathcal{E}[X_{MS}^2[k]] \tag{7-228}$$

观察可得：

$$\begin{aligned}\mathcal{E}[X_{MS}^2[k]] &= \mathcal{E}[X_{MS}[k]\,\mathcal{E}[X[k] \mid X[0],\cdots,X[k-1]]] \\ &= \mathcal{E}[\mathcal{E}[X_{MS}[k]X[k] \mid X[0],\cdots,X[k-1]]] = \mathcal{E}[X_{MS}[k]X(k]]\end{aligned} \tag{7-229}$$

式(7-227)得出结果为零。第二个等式成立是因为 $X_{MS}[k]$是 $\{X[0],\cdots,X[k-1]\}$ 的函数，可以被放到求期望算子中，在求期望运算中考虑到条件(在第 5 章中被称为因式分解特性)，则它是一个常数。最后一个表达式成立是因为嵌套期望的特性。

由之前的讨论，我们看到一个随机序列可被分解为一个可预测部分 $X_{MS}[k]$(在 MS 意义下)和新息 $V[k]$(MS 不可预测部分)的和，如下所示：

$$X[k] = X_{MS}[k] + V[k] \tag{7-230}$$

因此随机序列由一个基本“信号”和一个加性“噪声”序列组成。这个模型意味着我们可以将初始序列的任何一个部分看做系统的输出（系统不一定必须是线性的）。图 7-20 给出了可预测部分和新息是如何由 $X[k]$ 产生的框图。

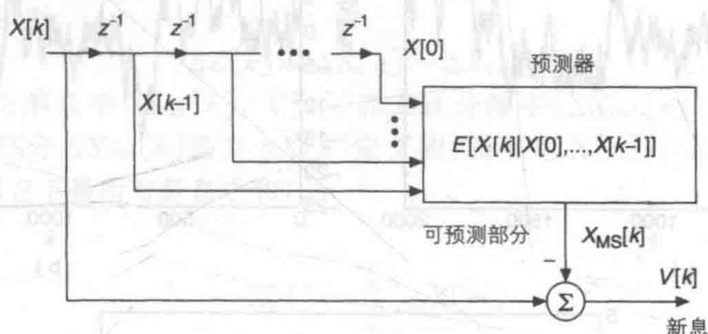


图 7-20 产生随机序列 $X[k]$ 的可预测部分 $X_{MS}[k]$ 和新息 $V[k]$ 的系统框图。系统 $\epsilon[X[k] \mid X[k-1],\cdots,X[0]]$ 是 $\{X[k-1],\cdots,X[0]\}$ 的一个函数（并不包括 $X[k]$）。因为假设序列是广义平稳的，系统可能是非线性的，但是是时不变的

例 7-26 对例 7-25 中的高斯-马尔可夫序列，新息序列为：

$$V[k] = X[k] - \rho X[k-1] - (1+\rho)\mu_X \tag{7-231}$$

可以直接得到这个结果：从 $X[k]$ 中减去从 $X[k-1]$ 预测的部分就是新息序列。预测的“可信度”由相关系数 ρ 大小决定。还减掉了一个包括均值的表达式，因为均值是已知和“可预测的”。特别注意到如果 $\rho=0$，则过程没有可预测部分，新息序列等于初始序列减去均值。图 7-21 证明了由 $X[k]$ 得到 $V[k]$ 的离散时间线性滤波器的实现过程以及逆过程（其中，为了方便，假设 $\mu_X=0$）。观察到产生 $V[k]$ 的滤波器是有限冲激响应（FIR），而产生 $X[k]$ 的是一个无限冲激响应（IIR）。IIR 滤波器有一个极点在 ρ，而且是稳定的，因为 $|\rho|\leqslant 1$（理论上，如果 $|\rho|=1$，这个滤波器只是边缘稳定的）。

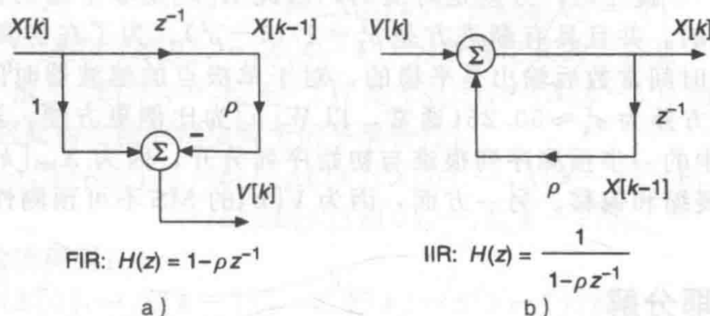


图 7-21 产生例 7-26 中序列的滤波器，其中 $\mu_X=0$。a) 新息序列 $V[k]$；b) 初始序列 $X[k]$ ◀

图 7-22 是高斯-马尔可夫过程的现实的例子以及它的新息。这个序列的产生过程如下：

$$X[k] = \rho X[k-1] + W[k] \tag{7-232}$$

其中 $W[k]$ 是一个零均值、单位方差的不相关高斯序列。这个表达式可写为以下非递归形式：

$$X[k] = \sum_{n=0}^{k} \rho^n W[k-n] \tag{7-233}$$

因此 $\mu_X=0$，且：

$$\sigma_X^2 = \sum_{n=0}^{k} \mathrm{var}[\rho^n W[k-n]] = \sum_{n=0}^{k} \rho^{2n} \mathrm{var}[W[k-n]]$$

$$= \sum_{n=0}^{k} \rho^{2n} = [1-\rho^{2(k+1)}]/(1-\rho^2) \tag{7-234}$$

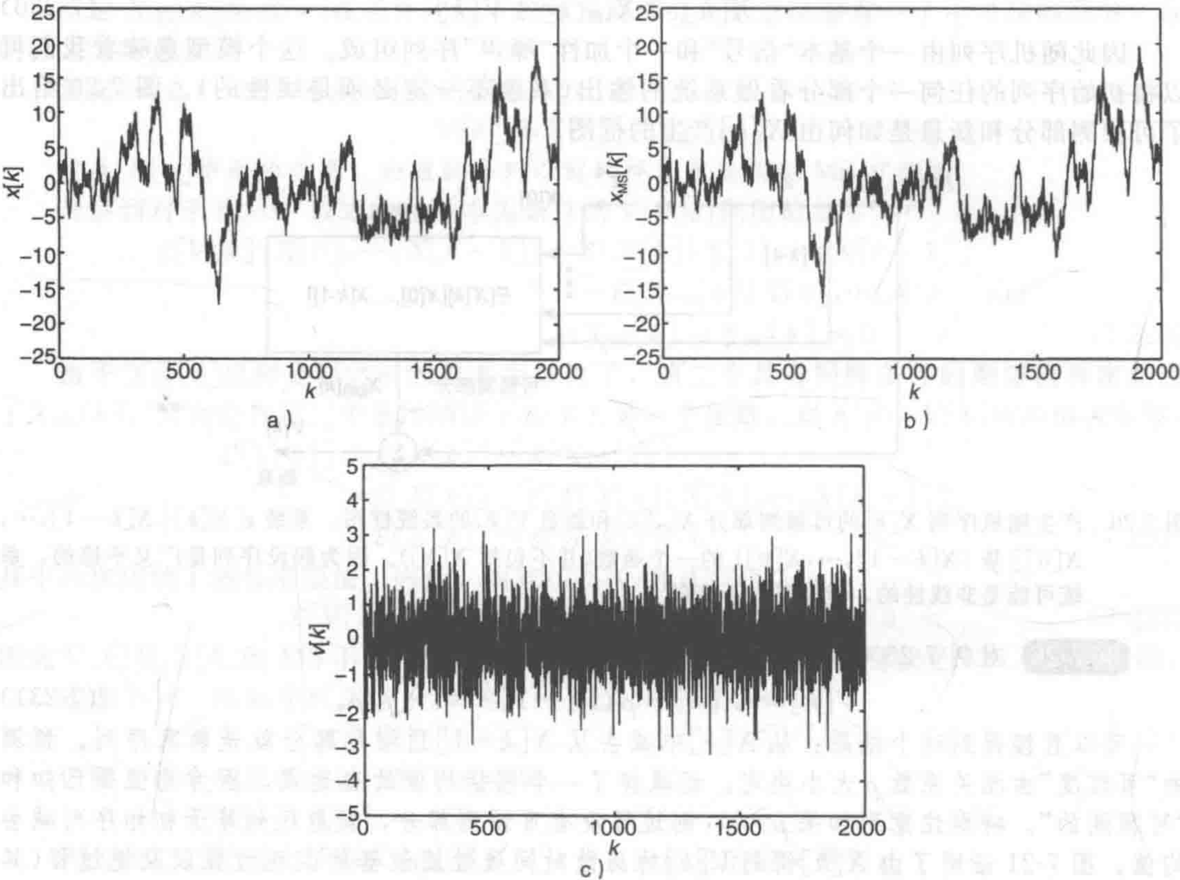


图 7-22 例 7-26 的现实，其中 $\rho=0.99$，$\mu_X=0$。a) 一个高斯-马尔可夫序列 $X[k]$的现实；b) 一步预测 $X_{MS}[k]$的现实；c) 新息 $V[k]$（注意新息幅值比例尺的变化）

由于序列从 $k=0$ 就开始，方差是时变的，因此 $X[k]$是非平稳的。但是，当 $k\to\infty$ 时，它是渐进平稳的，并且具有渐进方差 $\sigma_X^2=1/(1-\rho^2)$。为了在仿真中可用，我们假设在三到五个系统时间常数后输出是平稳的，对于单极点的滤波器时间常数大约为 $\tau\approx 1/(1-\rho)$。上式的方差为 $\sigma_X^2\approx 50.25$（通常，以 $W[k]$为比例更方便，这样 $X[k]$为单位方差）。观察到图中的一步预测序列很难与初始序列分开，因为 $X_{MS}[k]=0.99X[k-1]$ 只是 $X[k]$的一个展缩和偏移。另一方面，因为 $V[k]$的 MS 不可预测性，新息序列中有噪声的身影。

7.10 杜布-迈耶分解

接下来，我们给出随机序列 $X[k]$的另一种分解方法，它看起来像是新息分解法的一个改良。这种分解方法也是式(7-230)中的加法运算，但是分量序列具有与 $X_{MS}[k]$和 $V[k]$ 不同的特性，尽管它们彼此相关，稍后会证明这点。通过检测差值 $X[k]-X[k-1]$，我们发现被分解的 $X[k]$的一个分量是一个鞅序列。

定理 7-20（杜布-迈耶） $k\in\mathcal{Z}^+$，$\mathcal{E}[|X[k]|]<\infty$ 的随机序列可唯一表示为以下两个序列之和：

$$X[k]=X_{DM}[k]+E[k] \tag{7-235}$$

其中 $X_{DM}[k]$是 MS 可预测序列，$E[k]$是一个鞅。

注意到杜布-迈耶(DM)可预测部分 $X_{DM}[k]$与基于新息分解的 $X_{MS}[k]$是不同的。（尽管我们将式(7-235)看做 DM 分解，实际更精确的描述应该是对离散时间序列的杜布分解。杜布-迈耶是对应于连续时间过程的分解方法）

证明：我们先用以下的符号来定义增量序列：

$$\Delta X[k] \triangleq X[k]-X[k-1] \tag{7-236}$$

给定 $\{X[0],\cdots,X[k-1]\}$ 后增量序列的 DM 可预测部分是以下的条件均值：

$$\Delta X_{\mathrm{DM}}[k] \triangleq \mathcal{E}[\Delta X[k] \mid X[0],\cdots,X[k-1]] \tag{7-237}$$

由此可定义误差序列：

$$\Delta E[k] \triangleq \Delta X[k]-\Delta X_{\mathrm{DM}}[k] \tag{7-238}$$

注意到新息分解法中 $\{X_{\mathrm{MS}}[k],\ V[k]\}$ 和 DM 分解中 $\{\Delta X_{\mathrm{DM}}[k],\ \Delta E[k]\}$ 的相似性。因为 DM 可预测部分 $\Delta X_{\mathrm{DM}}[k]$ 是为 $\Delta X[k]$ 定义的(而不是 $X[k]$)，$\Delta E[k]$ 也与新息不同(稍后我们将证明它不是所有新息之和)。

观察得：

$$X[k]=\sum_{m=0}^{k-1} \Delta X[m] \tag{7-239}$$

因为当 $m=k-1$ 时增量中除了 $X[k]$ 外所有项相互抵消，假设 $X[-1]=0$。同样，以下基于增量的序列定义为：

$$X_{\mathrm{DM}}[k] \triangleq \sum_{m=0}^{k-1} \Delta X_{\mathrm{DM}}[m], E[k] \triangleq \sum_{m=0}^{k-1} \Delta E[m] \tag{7-240}$$

因此

$$X[k]=X_{\mathrm{DM}}[k]+E[k] \tag{7-241}$$

这就是我们需要的分解。我们强调 $X_{\mathrm{DM}}[k]$ 是由式(7-237)中的条件均值定义的 $X[k]$ 的可预测部分。假设 $X_{\mathrm{DM}}[0]=0$，因为对 $k<0$ 有 $X[k]=0$，则没有更早的样本能够预测 $X[0]$。因此，$E[0]=X[0]$，对 $k<0$，有 $E[k]=0$。

式(7-241)是 DM 分解：$X_{\mathrm{DM}}[k]$ 是序列的 MS 可预测部分(用 $\Delta X[k]$ 的表达式定义)，$E[k]$ 是一个鞅序列，我们现在来证明。观察式(7-238)，

$$\begin{aligned}\mathcal{E}[\Delta E[k] \mid [X[0],\cdots,X[k-1]] &= \mathcal{E}[\Delta X[k]-\Delta X_{\mathrm{DM}}[k] \mid X[0],\cdots,X[k-1]] \\ &= \Delta X_{\mathrm{DM}}[k]-\Delta X_{\mathrm{DM}}[k]=0\end{aligned} \tag{7-242}$$

其中用到了 $\Delta X_{\mathrm{DM}}[k]$ 是以 $\{X[0],\cdots,X[k-1]\}$ 为条件定义的，并且不会因为取期望值而发生改变的事实。由于 $\{E[0],\cdots,E[k-1]\}$ 是 $\{X[0],\cdots,X[k-1]\}$ 的函数，它们是基于一个航向 σ-域，因此由之前的期望算子的平滑特性：

$$\begin{aligned}\mathcal{E}[\Delta E[k] \mid X[0],\cdots,X[k-1]] &= \mathcal{E}[\mathcal{E}[\Delta E[k] \mid X[0],\cdots,X[k-1]] \mid E[0],\cdots,E[k-1]]] \\ &= \mathcal{E}[\Delta E[k] \mid E[0],\cdots,E[k-1]]\end{aligned} \tag{7-243}$$

代入增量中的项得到：

$$\begin{aligned}\mathcal{E}[\Delta E[k] \mid E[0],\cdots,E[k-1]] &= \mathcal{E}[E[k]-E[k-1] \mid E[0],\cdots,E[k-1]] \\ &= \mathcal{E}[E[k] \mid E[0],\cdots,E[k-1]]-E[k-1]\end{aligned} \tag{7-244}$$

从式(7-242)和式(7-243)得到等式左边为零，我们最终可得：

$$\mathcal{E}[E[k] \mid E[0],\cdots,E[k-1]]=E[k-1] \tag{7-245}$$

证明 $E[k]$ 是一个鞅。

接下来，我们推导新息和 DM 分解的相互关系。由 $\Delta X_{\mathrm{DM}}[k]$ 和 $\Delta E[k]$ 的定义，我们得到：

$$\begin{aligned}\Delta X_{\mathrm{DM}}[k] &= \mathcal{E}[X[k] \mid X[0],\cdots,X[k-1]]-\mathcal{E}[X[k-1] \mid X[0],\cdots,X[k-1]] \\ &= X_{\mathrm{MS}}[k]-X[k-1]\end{aligned} \tag{7-246}$$

其中 $X_{\mathrm{MS}}[k]$ 是新息分解中的 MS 可预测部分。将这个结果与式(7-238)联合可得：

$$\begin{aligned}\Delta E[k] &= X[k]-X[k-1]-(X_{\mathrm{MS}}[k]-X[k-1]) \\ &= X[k]-X_{\mathrm{MS}}[k]=V[k]\end{aligned} \tag{7-247}$$

这是新息。因此，DM 分解中的鞅分量是所有新息样本的和：

$$E[k]=\sum_{m=0}^{k}\Delta E[m]=\sum_{m=0}^{k}V[m] \tag{7-248}$$

这两类序列分解的各种分量在图 7-23 中进行了总结。

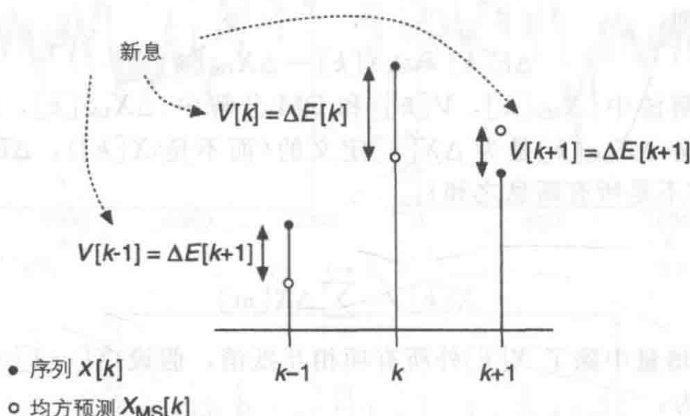


图 7-23　$X[k]$的 MS 预测样本 $X_{\mathrm{MS}}[k]$，新息 $V[k]$以及 DM 分解的鞅分量 $E[k]$

由于 $E[k]$是所有新息之和，我们可以用一种比以前的证明更直接的方法来证明它是一个鞅。代入新息的和：

$$\mathcal{E}[E[k]\,|\,E[0],\cdots,E(k-1]]=\sum_{m=0}^{k}\mathcal{E}[V[m]\,|\,E[0],\cdots,E[k-1]] \tag{7-249}$$

将第 k 项分开得到：

$$\mathcal{E}[E[k]\,|\,E[0],\cdots,E[k-1]]=\sum_{m=0}^{k-1}\mathcal{E}[V[m]\,|\,E[0],\cdots,E[k-1]]$$
$$+\mathcal{E}[V[k]\,|\,E[0],\cdots,E[k-1]]=E[k-1] \tag{7-250}$$

等式右边第一项是希望的结果 $E[k-1]$，因为它包括在条件中；第二项为零，因为$\{E[0],\cdots,E[k-1]\}$只依赖于新息样本$\{V[0],\cdots,V[k-1]\}$到时刻 $k-1$，它是 MS 不可预测的，均值为零。

最后，通过以前的结果，我们来看 DM 分解到新息分解的转换是如何实现的：

$$X[k]=X_{\mathrm{DM}}[k]+E[k]=\sum_{m=0}^{k}\Delta X_{\mathrm{DM}}[m]+\sum_{m=0}^{k}V[m]$$
$$=\sum_{m=0}^{k}X_{\mathrm{MS}}[m]-\sum_{m=0}^{k-1}X[m]+\sum_{m=0}^{k}V[m] \tag{7-251}$$

从第一个和第三个求和项中分出 $m=k$ 项，得到：

$$X[k]=X_{\mathrm{MS}}[k]+V[k]+\sum_{m=0}^{k-1}(X_{\mathrm{MS}}[m]-X[m]+V[m]) \tag{7-252}$$

这就是我们希望的结果，因为 $V[m]\triangleq X[m]-X_{\mathrm{MS}}[m]$，因此最后一个求和项为零。

例 7-27　继续分析例 7-26，我们计算 $\mu_X=0$ 的高斯-马尔可夫序列的 DM 分解。可预测部分为：

$$X_{\mathrm{DM}}[k]=\sum_{m=0}^{k}X_{\mathrm{MS}}[m]-\sum_{m=0}^{k-1}X[m]$$
$$=\sum_{m=1}^{k}\rho X[m-1]-\sum_{m=0}^{k-1}X[m]=\sum_{m=0}^{k-1}(\rho-1)X[m] \tag{7-253}$$

鞅序列为：

$$E[k]=X[k]-\sum_{m=0}^{k-1}(\rho-1)X[m] \tag{7-254}$$

也可将式(7-231)中新息($\mu_X=0$)求和：

$$E[k]=\sum_{m=0}^{k}(X[m]-\rho X[m-1])$$
$$=X[k]+\sum_{m=0}^{k-1}X[m]-\rho\sum_{m=1}^{k}X[m-1] \tag{7-255}$$

改变最后一个求和项中的变量可得到式(7-254)中的结果。图 7-24 给出了 DM 分解的框图，显示了 DM 可预测部分和鞅分量是怎样产生的。图 7-25 给出了 $X_{DM}[k]$ 和 $E[k]$ 的现实的例子，与图 7-20 中的原始序列以及它的新息进行比较。

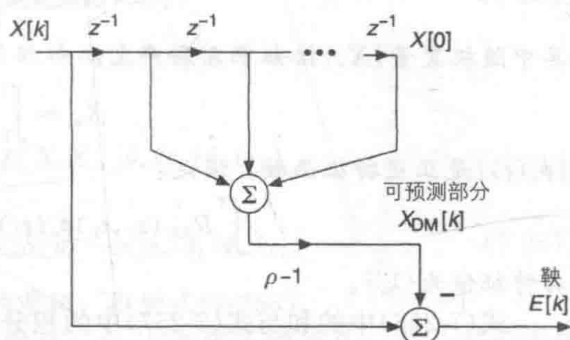


图 7-24 产生例 7-27 中 $\mu_X=0$ 的序列的滤波器。a)可预测部分 $X_{DM}[k]$；b)鞅序列 $E[k]$

图 7-25 例 7-27 中 $\rho=0.99$，$\mu_X=0$ 时的现实。a) 高斯-马尔可夫序列 $X[k]$的现实；b) 杜布-迈耶可预测部分 $X_{DM}[k]$；(c) 杜布-迈耶鞅部分 $E[k]$

7.11 卡胡内-列维展开

在最后一节，我们回到对连续时间随机过程的讨论中，描述一个由正交函数的加权和来实现的分解，它被称为卡胡内-列维展开(KLE)。

定理 7-21(卡胡内-列维) 设 $X(t)$是一个零均值随机过程，自相关函数为定义在有限

区间 $\mathcal{T}=[0,T]$ 上的函数 $R_{XX}(t_1,t_2)$。它可以做如下分解：

$$X(t)=\sum_{n=1}^{\infty}X_n\phi_n(t) \tag{7-256}$$

其中随机变量 $\{X_n\}$ 按如下方法产生：

$$X_n=\int_0^T X(t)\phi_n(t)\mathrm{d}t \tag{7-257}$$

$\{\phi_n(t)\}$ 是正交特征函数，满足：

$$\int_0^T R_{XX}(t_1,t_2)\phi_n(t_2)\mathrm{d}t_2=\lambda_n\phi_n(t_1),t_1\in\mathcal{T} \tag{7-258}$$

其特征值为 $\{\lambda_n\}$。

式(7-256)中的和与式(7-257)中的积分都在 MS 意义下定义，而最后一个积分项是基于确定函数乘积。式(7-258)中的自相关函数也被称为积分方程的核。

证明：设式(7-256)中 N 个元素的和为有限值，则它为一个序列。对 MS 收敛，我们需要：

$$\lim_{N\to\infty}\mathcal{E}\left[\left(X(t)-\sum_{n=1}^{N}X_n\phi_n(t)\right)^2\right]=0 \tag{7-259}$$

将期望运算展开得到：

$$\mathcal{E}\left[\left(X(t)-\sum_{n=1}^{N}X_n\phi_n(t)\right)^2\right]=\mathcal{E}[X^2(t)]-2\sum_{n=1}^{N}\mathcal{E}[X(t)X_n]\phi_n(t)$$
$$+\sum_{n=1}^{N}\sum_{m=1}^{N}\mathcal{E}[X_nX_m]\phi_n(t)\phi_m(t) \tag{7-260}$$

将式(7-257)代入等式右边第二项：

$$\mathcal{E}[X(t)X_n]=\int_0^T\mathcal{E}[X(t)X(t_2)]\phi_n(t_2)\mathrm{d}t_2=\int_0^T R_{XX}(t,t_2)\phi_n(t_2)\mathrm{d}t_2=\lambda_n\phi_n(t) \tag{7-261}$$

其中用到了式(7-258)。同样对式(7-260)中的最后一项：

$$\begin{aligned}\mathcal{E}[X_nX_m]&=\int_0^T\int_0^T\mathcal{E}[X(t_1)X(t_2)]\phi_n(t_1)\phi_m(t_2)\mathrm{d}t_1\mathrm{d}t_2\\&=\int_0^T\int_0^T R_{XX}(t_1,t_2)\phi_n(t_1)\phi_m(t_2)\mathrm{d}t_1\mathrm{d}t_2\\&=\int_0^T\lambda_m\phi_m(t_1)\phi_n(t_1)\mathrm{d}t_1=\lambda_m\delta[m-n]\end{aligned} \tag{7-262}$$

其中我们用到了特征函数的正交特性：

$$\int_0^T\phi_n(t)\phi_m(t)\mathrm{d}t=\delta[m-n] \tag{7-263}$$

这是克罗内克 δ 函数。证明了 KLE 中的随机变量是正交的。将以上所有结果联合，式(7-260)变为：

$$\begin{aligned}\mathcal{E}\left[\left(X(t)-\sum_{n=1}^{N}X_n\phi_n(t)\right)^2\right]&=R_{XX}(t,t)-2\sum_{n=1}^{N}\lambda_n\phi_n^2(t)+\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_m\delta[m-n]\phi_n(t)\phi_m(t)\\&=R_{XX}(t,t)-\sum_{n=1}^{N}\lambda_n\phi_n^2(t)\end{aligned} \tag{7-264}$$

其中 δ 函数将二重求和减为一个，然后又被中间一项减去。设 $N\to\infty$：

$$R_{XX}(t,t)-\lim_{N\to\infty}\sum_{n=1}^{N}\lambda_n\phi_n^2(t)=0 \tag{7-265}$$

得到式(7-259)。最后的结果是由以下的 Mercer 定理导出的(其中 $t=t_1=t_2$)。

注意，$R_{XX}(t,t)\neq R_{XX}(0)$，除非 $X(t)$ 是广义平稳的：对非平稳过程 $X(t)$，这两个时间变量必须明确的给出。

定理 7-22(Mercer) 随机过程 $X(t)$的自相关函数可做如下分解：

$$R_{XX}(t_1,t_2)=\sum_{n=1}^{\infty}\lambda_n\phi_n(t_1)\phi_n(t_2) \tag{7-266}$$

其中$\{\lambda_n\}$和$\{\phi_n(t)\}$分别是之前定义的 KLE 的特征值和特征函数。

证明：将式(7-256)代入得到：

$$\begin{aligned}\mathcal{E}[X(t_1)X(t_2)]&=\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\mathcal{E}[X_nX_m]\phi_n(t_1)\phi_m(t_2)\\&=\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\lambda_m\delta[m-n]\phi_n(t_1)\phi_m(t_2)\end{aligned} \tag{7-267}$$

用式(7-262)中的δ函数的结论消去对m的求和，得到式(7-266)。

观察可知 KLE 与矩阵的特征分解类似(见附录 G)。与矩阵不同的是，式(7-258)中的自相关函数决定了特征函数的类型。图 7-26给出了分解的框图。在第 10 章，我们会发现这个结构与匹配滤波接收机类似。

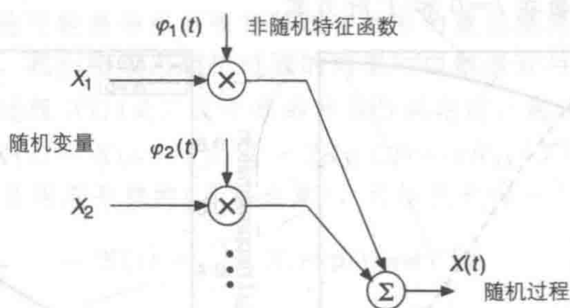


图 7-26 随机过程 $X(t)$用由随机变量$\{X_n\}$加权的特征函数$\{\phi_n(t)\}$表示的 KLE($n\in\mathcal{N}$)

例 7-28 考虑一个自相关函数为 $R_{XX}(t_1,t_2)=c\min(t_1,t_2)$ 的维纳过程，其中 $c=\alpha\varepsilon^2$ (见表 6-4)。假设 $X(t)$在 $\mathcal{T}=[0,T]$ 区间定义，式(7-258)给出：

$$c\int_0^T\min(t_1,t_2)\phi_n(t_2)\mathrm{d}t_2=c\int_0^{t_1}t_2\phi_n(t_2)\mathrm{d}t_2+c\int_{t_1}^T t_1\phi_n(t_2)\mathrm{d}t_2=\lambda_n\phi_n(t_1) \tag{7-268}$$

其中 $t_1\in\mathcal{T}$。用解微积分方程的方法(比如在二阶 RLC 电路中用到的方法)，对这个表达式微分得到函数积分时明确的表达式。由链微分法和莱布尼茨法则(见附录 E)：

$$\frac{\partial}{\partial t_1}\int_0^{t_1}t_2\phi_n(t_2)\mathrm{d}t_2=t_1\phi_n(t_1) \tag{7-269}$$

$$\frac{\partial}{\partial t_1}\int_{t_1}^T t_1\phi_n(t_2)\mathrm{d}t_2=-t_1\phi_n(t_1)+\int_{t_1}^T\phi_n(t_2)\mathrm{d}t_2 \tag{7-270}$$

对式(7-268)微分，代入这些表达式得到一个简化的积分方程，它定义了特征函数：

$$c\int_{t_1}^T\phi_n(t_2)\mathrm{d}t_2=\lambda_n\frac{\partial}{\partial t_1}\phi_n(t_1) \tag{7-271}$$

很容易再对 t_1 进行二次微分，得到二阶 DE：

$$\frac{\mathrm{d}^2}{\mathrm{d}t_1^2}\phi_n(t_1)=-(c/\lambda_n)\phi_n(t_1) \tag{7-272}$$

单边拉普拉斯变换为：

$$s^2\Phi_n(s)-s\phi_n(0)-\phi'_n(0)=-(c/\lambda_n)\Phi_n(s) \tag{7-273}$$

假设零初始条件，特征方程为：

$$s^2+c/\lambda_n=0 \tag{7-274}$$

如果自相关矩阵是正定的(PD)，则 $\lambda_n>0$(见习题 7-28)，因此特征方程的根为 $s=\mathrm{j}\sqrt{c/\lambda_n}(c>0)$。通解为：

$$\phi_n(t) = a_1\sin(t\sqrt{c/\lambda_n}) + a_2\cos(t\sqrt{c/\lambda_n}) \tag{7-275}$$

由 $\phi_n(0)=0$，得到 $a_2=0$。λ_n 的解由以下在 $t=T$ 时刻的条件得到：

$$\phi'_n(T) = 0 \Rightarrow a_1\sqrt{c/\lambda_n}\cos(T\sqrt{c/\lambda_n}) = 0 \tag{7-276}$$

这意味着 $T\sqrt{c/\lambda}$ 为 $\pi/2$ 的奇数倍。对于 $n\in\mathcal{N}$：

$$T\sqrt{c/\lambda_n} = (2n-1)\pi/2 \Rightarrow \lambda_n = 4cT^2/(2n-1)^2\pi^2 = cT^2/(n-1/2)^2\pi^2 \tag{7-277}$$

将这个结果代入式(7-275)得到：

$$\phi_n(t) = a_1\sin((n-1/2)\pi t/T) \tag{7-278}$$

为了特征函数具有单位能量，最直接的方法就是证明 $a_1=\sqrt{2/T}$。图 7-27a 中画出了前三个特征函数。注意到对任意 $T>0$，在 $[0,T]$ 上得到同样的波形，尽管振幅随着 T 增长而衰减，因此函数保持单位能量。在图 7-27b 中，我们用 Mercer 准则给出了求和项中 $N=1$ 和 6 时 $R_{XX}(t,t)=\mathcal{E}[X^2(t)]$ 的近似值。对 $c=1$，注意到 $R_{XX}(t,t)=\min(t,t)=t$ 是线性的，则它的近似值是非常精确的，因为求和项中只有几项。因为和式中每个特征函数的边界条件，曲线的导数在 $t=0$ 和 T 时为零。

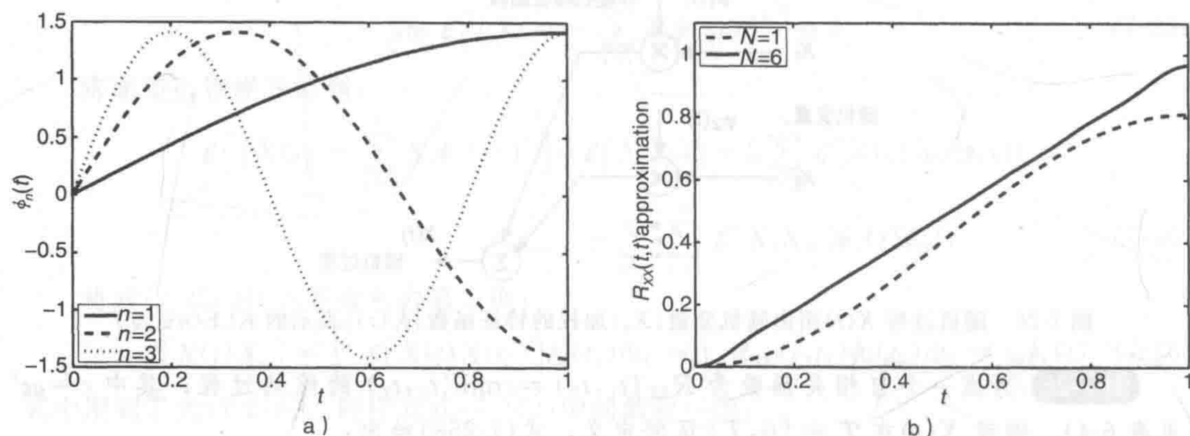


图 7-27　$T=1$ 时例 7-28 的 KLE。a) $R_{XX}(t_1,t_2)$ 的特征函数的例子；b) 用 Mercer 准则中有限项的和的 $R_{XX}(t,t)$ 近似 ◀

通常很难导出特征函数 $\{\phi_n(t)\}$ 的表达式，常需要采用数值解法。在例 7-29 中，我们证明即使对一个相对简单的随机过程这也是比较困难的任务。

例 7-29　考虑以下定义在 $[0,T]$ 上的随机过程：

$$X(t) = \sum_{m=1}^{M} A_m g_m(t) \tag{7-279}$$

其中 $\{g_m(t)\}$ 是确定函数。$X(t)$ 中唯一的随机量是随机变量 $\{A_m\}$，我们假设其均值为零，具有有限方差 $\{\sigma_m^2\}$。当随机变量独立时，自相关函数为：

$$R_{XX}(t_1,t_2) = \sum_{m=1}^{M} \sigma_m^2 g_m(t_1) g_m(t_2) \tag{7-280}$$

由积分方程，

$$\sum_{m=1}^{M} \sigma_m^2 \int_0^T g_m(t_1) g_m(t_2) \phi_n(t_1)\,dt_2 = \lambda_n \phi_n(t_1) \tag{7-281}$$

我们得到特征函数依赖于 $\{g_m(t)\}$ 的特定形式。推导中关键步骤是将 $g_m(t_1)$ 从积分中提出来，然后用等式左边的其他项关于 t_1 独立这个事实得到：

$$\sum_{m=1}^{M} \sigma_m^2 g_m(t_1) \int_0^T g_m(t_2) \phi_n(t_1)\,dt_2 = \sum_{m=1}^{M} b_{m,n} g_m(t_1) \tag{7-282}$$

其中

$$b_{m,n} \triangleq \sigma_m^2 \int_0^T g_m(t_2)\phi_n(t)\mathrm{d}t_2 \tag{7-283}$$

尽管需要特征函数来给出 $\{b_{m,n}\}$ 的特定值，我们看到通过将式(7-282)与式(7-281)的右边画等号，得到特征函数是 $\{g_m(t)\}$ 的线性组合：

$$\phi_n(t) = \sum_{m=1}^{M} c_{m,n} g_m(t) \tag{7-284}$$

其中 $c_{m,n} \triangleq b_{m,n}/\lambda_n$。如果要继续运算，还需要关于 $\{g_m(t)\}$ 的更多信息。例如，我们可以通过适当的调整方差来使它们具有单位能量 $\int_0^T g_m^2(t)\mathrm{d}t = 1$。更重要的是，我们需要确定互相关 $\int_0^T g_m(t)g_n(t)\mathrm{d}t$，它会直接影响 $X(t)$ 的现实和式(7-280)中自相关矩阵的特性。通常对这个例子不可能推导出 $\{c_{m,n}\}$ 闭合形式的表达式；实际上，对 $M=2$，解不是唯一的(Ludeman，2003)。第 10 章的数字通信中，我们将给出一个类似于 KLE 的扩展，称做 Gram-Schmidt 正交化，可用于确定性信号，类似于本例中的 $\{g_n(t)\}$ 函数。正交化使发射波形的表示更方便，由此可能推导出在噪声信道中信号的最优检测器。◀

在最后一个例题中，我们将给出随机过程的傅里叶级数展开与 KLE 的区别。

例 7-30 假设随机过程 $X(t)$ 是广义平稳的和 MS 周期的，则

$$\mathcal{E}[(X(t)-X(t-T))^2] = 2R_{XX}(0) - 2R_{XX}(T) = 0 \tag{7-285}$$

对这种情况，$X(t)$ 是周期平稳的(见第 6 章)，可以展开为一个傅里叶级数

$$X(t) = \sum_{n=-\infty}^{\infty} X_n \exp(\mathrm{j}2\pi nt/T) \tag{7-286}$$

其中 $\{X_n\}$ 是随机变量：

$$X_n = \frac{1}{T}\int_0^T X(t)\exp(-\mathrm{j}2\pi nt/T)\mathrm{d}t \tag{7-287}$$

以上的求和与积分都是在 MS 意义下定义的。注意到这个展开式与 KLE 类比有以下区别：(ⅰ) $\{\phi_n(t)\}$ 被复指数取代了；(ⅱ)求和扩展到了对所有 $n \in \mathcal{Z}$。KLE 假设 $X(t)$ 只在 $[0,T]$ 上存在。如果我们设想这个过程以 T 为周期重复，则式(7-285)成立就能推导出其傅里叶级数表达式。由于 $R_{XX}(\tau)$ 是以 T 为周期的(见习题 7-33)，它也能用傅里叶级数展开。

$$R_{XX}(\tau) = \sum_{n=-\infty}^{\infty} r_n \exp(\mathrm{j}2\pi n\tau/T) \tag{7-288}$$

傅里叶级数的系数：

$$r_n = \frac{1}{T}\int_0^T R_{XX}(\tau)\exp(-\mathrm{j}2\pi n\tau/T)\mathrm{d}\tau \tag{7-289}$$

这个积分方程与 KLE 的式(7-258)类似。用类似于 KLE 中的导数方法，可以证明 $X(t)$ 的傅里叶级数在 MS 意义下存在(见习题 7-34)。◀

习题

随机收敛

7-1 确定以下哪个随机序列收敛：(ⅰ)必然收敛，(ⅱ)几乎必然。

(a) $X_1[k,\zeta] = \zeta(k-1)/k$，$\mathcal{S}=[0,1]$

(b) $X_2[k,\zeta] = (1/\zeta)(k-1)/k$，$\mathcal{S}=[0,\infty)$

(c) $X_3[k,\zeta] = \sin(2\pi\zeta k)$，$\mathcal{S}=[-1,1]$

7-2 确定以下哪个随机序列依 MS 收敛。

(a) $X_1[k,\zeta] = \zeta/k$，$\mathcal{S}=[0,1]$.

(b) $X_2[k,\zeta] = \zeta(k-1)/k$，$\mathcal{S}=[0,1]$

(c) $X_3[k,\zeta] = \exp(\zeta k)$，$\mathcal{S}=[-1,1]$

7-3 假设 $X[k]$ 依概率收敛于常数 a，$Y[k]$ 依概率收敛于常数 b。由定义证明 $X[k]+Y[k]$ 依概率收敛于 $a+b$。

7-4 证明对于 $m>n$，L^m 收敛意味着 L^n 收敛。

7-5 设 $Y[k] = p^k \prod_{n=1}^{k} X[k]$，其中 $X[k]$ 是一个独立同分布伯努利序列，参数为 p。确定这个序列是否对 $X[k]$ 的以下输出几乎必然收敛：

(a) $\{-1,1\}$；(b) $\{0,1\}$。

7-6 对 $X[k]\xrightarrow{\text{m.s.}}X$ 和 $Y[k]\xrightarrow{\text{m.s.}}Y$，证明以下结论：

(a) $X[k]+Y[k]\xrightarrow{\text{m.s.}}X+Y$；

(b) $\lim\limits_{k\to\infty}\mathcal{E}[X[k]Y[k]]=\mathcal{E}[XY]$。

大数定理

7-7 证明对一个输出为 $\{0,1\}$，参数为 p 的独立同分布伯努利序列弱大数定理成立。

7-8 设 $X[k]$是一个独立同分布序列，弱大数定理对其成立。确定弱大数定理也可应用于$Y[k]=g(X[k])$，其中 $g(\cdot)$为一个无记忆函数。

7-9 假设弱大数定理的条件放松到 $X[k]$不是一个独立的序列。对于自协方差函数 $C_{XX}[m]=\exp(-|m|)(m\in Z)$，确定式(7-42)中的特性是否仍有效。

中心极限定理

7-10 设 $X[k]$为一个独立同分布指数随机变量序列，参数 $\lambda=1$，定义求和项 $Y[k]\triangleq(1/k)\sum_{n=1}^{k}X[n]$计算以下概率，用(i)$Y[k]$的确定分布(ii)基于 CLT 的一个高斯近似。

(a) $P(Y[5]>2)$；(b) $P(1<Y[10]\leqslant 2)$。

7-11 重复习题 7-10 的问题，对参数 $p=3/4$ 的伯努利随机变量，计算以下概率。

(a) $P(Y[3]<2)$；(b) $P(1\leqslant Y[4]\leqslant 2)$。

7-12 对二元对称信道(BSC)(见第 2 章)，设一比特的误差概率为 $p_e=0.01$。如果 500bit 通过这个信道传输，用 CLT 来计算大于 5 个错误的概率。

7-13 推导在 CLT 的证明中用到的式(7-57)中的变换结论。

7-14 考虑一个参数 $\alpha=10$ 的泊松随机变量 Y。认识到这个分布可通过将 10 个独立同分布随机变量求和来得到，我们可以应用 CLT 和式(4-17)中的结论，用高斯近似求 $P(Y>8)$。

连续、导数和积分

7-15 设 $X(t)=u(t)-u(t-T)$，其中 $u(t)$是单位阶跃函数，T 是在 $[1,2]$ 上均匀分布的随机变量。确定 $X(t)$是否是 MS 连续的。

7-16 假设对$\mathcal{T}=\mathcal{R}$，有 $X(t)=\zeta\cos(2\pi f_0 t)$，其中 f_0 是一个常数频率，$\zeta\in\mathcal{S}=[-1,1]$ 是均匀分布。确定 $X(t)$是否：(a)依概率 1 连续；(b)MS 连续。

7-17 证明 MS 导数 $X'(t)$与广义平稳过程 $X(t)$不相关。

7-18 证明如果 $X(t)$对每个 $t\in\mathcal{T}$都是 MS 连续的，则 $R_{XX}(t_1,t_2)$ 对所有 t_1，$t_2\in\mathcal{T}$是连续函数。

7-19 当 $X(t)$广义平稳时，推导 MS 导数在式(7-130)和式(7-131)中的相关结果。

7-20 确定以下的随机过程是否具有 MS 导数和 MS 积分：$Y(t)$自相关函数

$$R_{YY}(\tau)=1-\tau^2,\quad |\tau|\in[0,1]$$

7-21 确定习题 7-15 中的 $X(t)$是否具有 MS 导数和 MS 积分。

7-22 计算例 7-24 中 $R_{XX}(t_1,t_2)$ 的导数。

7-23 与例 7-22 中方法类似，推导以下一阶 DE 的 $R_{XY}(t_1,t_2)$ 和 $R_{XX}(t_1,t_2)$ 的表达式：

$$a_1\frac{\mathrm{d}}{\mathrm{d}t}X(t)+a_0X(t)=Y(t)\tag{7-290}$$

对 $t>0$，$R_{YY}(\tau)=\delta(\tau)$

随机序列和分解

7-24 考虑以下一阶差分方程：

$$X[k]+0.5X[k-1]=Y[k]\tag{7-291}$$

设 $R_{YY}[m]=\delta[k]$，则 $Y[k]$是零均值不相关序列。假设差分方程具有零初始条件，用 z 变换求以下表达式：(a) $R_{XY}[k_1,k_2]$；(b) $R_{XX}[k_1,k_2]$。

7-25 设零均值马尔可夫序列 $X[k]$具有以下条件 pdf：

$$f_{X[k]\mid X[k-1]}(x[k]\mid x[k-1])=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x[k]-ax^2[k-1])^2/2\sigma^2)\tag{7-292}$$

其中 $a>0$。求：

(a) MS 可预测部分 $X_{\text{MS}}[k]=\varepsilon[X[k]\mid X[0],\cdots,X[k-1]]$

(b) DM 可预测部分 $\Delta X_{\text{DM}}[k]=\varepsilon[X[k]-X[k-1]\mid X[0],\cdots,X[k-1]]$

7-26 对式(6-160)中定义的二项式计数序列 $X[k]$，求 DM 可预测部分 $\Delta X_{\text{DM}}[k]$并给出分解 $X[k]=X_{\text{DM}}[k]+E[k]$中的各分量值。

卡胡内-列维展开

7-27 设 $X(t)$ 的自相关函数为 $R_{XX}(t_1,t_2)=\sigma^2\delta(t_2-t_1)$。证明任意正交特征函数集可用于 $X(t)$的 KLE。

7-28 式(7-258)中积分方程的核是 PD，如果

$$\int_0^{\infty}\int_0^{\infty}g(t_1)R_{XX}(t_1,t_2)g(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2>0\tag{7-293}$$

对任意在 $[0,T]$ 上的 $g(t)\neq 0$ 成立。用此结果来证明 KLE 的特征值$\{\lambda_n\}$为正数。

7-29 用 Mercer 定理求以下积分式用特征值来表示的表达式

$$\int_0^{T}R_{XX}(t,t)\,\mathrm{d}t,\int_0^{T}\int_0^{T}R_{XX}^2(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\tag{7-294}$$

7-30 对以下自相关函数：

$$R_{XX}(t_1,t_2)=2+\sin(2\pi t_1/T)\sin(2\pi t_2/T)\tag{7-295}$$

用 Mercer 定理求特征值 $\{\lambda_n\}$ 和特征函数 $\{\phi_n(t)\}$。

7-31 求定义在 $[0, T]$ 上的广义平稳随机过程 $X(t)$ 的 KLE，$X(t)$ 的自相关函数为：

$$R_{XX}(\tau) = T - |\tau| \quad (7\text{-}296)$$

7-32 重复习题 7-31 的问题，随机过程为：

$$X(t) = a\cos(2\pi f_c t + \Phi) \quad (7\text{-}297)$$

其中 $\{a, f_c\}$ 是非随机的，Φ 在 $[-\pi, \pi]$ 上均匀分布。

7-33 设 $X(t)$ 为一个零均值广义平稳的随机过程，它是 MS 周期的，周期为 T。证明自相关函数是具有同周期 T 的周期量。

7-34 (a)证明一个 MS 周期随机过程的傅里叶级数展开在 MS 意义下存在。KLE 中用过的方法也可以用在这里，除了 $\{X_n\}$ 是复数，因为是复指数。

(b)证明 $\mathcal{E}[X_n X_m^*] = r_n \delta[m-n]$。

仿真作业

7-35 设 $X[k]$ 是一个独立同分布随机序列，具有贝塔分布，参数为 $\alpha = \beta = 1/2$（对应反正弦分布）。用 MATLAB 计算式(7-55)中 $Y[k]$ 的 pdf 曲线，证明随着 k 的增长，它趋近高斯 pdf。由于本题中样本均值不具有闭合形式的 pdf，对每个 k 用卷积，包括式(7-55)中平移和展缩。

7-36 用 MATLAB 生成例 7-27 中类似于图 7-25 中 DM 分解的曲线图。设 $\mu_X = 0$，画 $\rho = \pm 0.9$ 时可预测部分 $X_{DM}[k]$ 和缺 $E[k]$ 的现实。

7-37 用 Mercer 定理的有限和，画维纳过程的自相关函数 $R_{XX}(t_1, t_2)$ 的近似图形。用 meshgrid 和 mesh 画这个函数的三维图，说明它们接近图 6-27 中的真实结果。注意图 7-27中显示的例 7-28 中的自相关函数的近似图只沿着 $t_1 = t_2 = t$ 这条直线。

进一步阅读

本章的材料在第 3 章结束部分的几本参考书中有进一步地探讨。在以下参考书中可以找到对一些特定问题的讨论，随机运算：Gardner(1990)，Larson 和 Shubert(1979)，Parzen(1962)、Stark 和 Woods(2002)。杜布-迈耶分解：Doob(1953)，Larson 和 Shubert(1979)。卡胡内-列维展开：Helstrom(1995)，Ludeman(2003)，Schonhoff 和 Giordana(2006)，Van Trees(1968)。

第8章 系统、噪声和谱估计

8.1 引言

本章主要考虑加性噪声消除、通信信道失真补偿等为特定滤波目的而设计的系统。一般而言我们采用线性时不变(Linear Time Invariant，LTI)系统，它们可通过冲激响应函数 $h(t)$ 或相应的传输函数 $H(\omega)$ 描述。这些系统也可以用常系数线性微分方程(Differential Equation，DE)表示，因此它们由信号的微分或积分运算构成。第 7 章研究了均方(Mean Square，MS)意义下随机过程的微分、积分的定义，进而得到了自相关函数 $R_{XX}(t_1,t_2)$。表 8-1 总结了均方的连续性、导数和积分；也包括广义平稳过程，即自相关函数 $R_{XX}(\tau)$ 仅依赖时间差 $\tau \triangleq t_2-t_1$ 的结果。本章主要研究广义平稳过程，对于所有 $t\in\mathcal{T}$，它们满足：

$$\mu_X(t)=\mu_X,\quad R_{XX}(t,t+\tau)=R_{XX}(\tau) \tag{8-1}$$

表 8-1 随机过程的均方结论

连续性：	$R_{XX}(t_1,t_2)$ 在 $t_1=t_2=t$ 时连续，或 $R_{XX}(\tau)$ 在 $\tau=0$ 时连续
可导：	$\partial^2 R_{XX}(t_1,t_2)/\partial t_1\partial t_2$ 在 $t_1=t_2=t$ 时存在，或 $\partial^2 R_{XX}(\tau)/\partial\tau^2$ 在 $\tau=0$ 时存在
可积：	$\int_{t_0}^{t}\int_{t_0}^{t} R_{XX}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2$ 存在，或 $2\int_0^{t-t_0}(t-t_0-\tau)R_{XX}(\tau)\mathrm{d}\tau$ 存在

如何分析随机过程通过 LTI 系统时发生的变化是本章研究的主要内容，包括时域和频域两方面。由于离散时间系统的卷积运算使用求和而非积分，本章还研究了针对离散时间系统的相应分析方法。当然，如果离散时间系统冲激响应用狄拉克函数表示，其卷积和也可通过积分计算。

本章最重要的内容是功率谱密度(Power Spectral Density，PSD) $S_{XX}(\omega)$，它是自相关函数 $R_{XX}(\tau)$ 的傅里叶变换。当 $X(t)$ 仅由式(8-1)的属性表示时，我们得到了随机过程的部分性质。在大部分应用中，我们并不关心随机过程的具体分布或高阶矩等统计量。式(8-1)中的描述对于大多数随机过程和应用而言极为有用。第 1 章分析了确定性信号和 LTI 系统的频域性质。对于确定性信号 $x(t)$，通过冲激响应为 $h(t)$ 的线性时不变系统的输出为卷积 $y(t)=h(t)*x(t)$，对应到频域上，$Y(\omega)=H(\omega)X(\omega)$，其中 $H(\omega)$ 是系统的传递函数。对于广义平稳随机过程，本章将说明与之对应的频域结果是 $S_{YY}(\omega)=|H(\omega)|^2\ S_{XX}(\omega)$，如图 8-1 所示。出现相同结果的原因与 LTI 系统一样，对于任意 $\tau\in\mathcal{R}$，$R_{XX}(\tau)$ 的特征函数也是复指数函数。我们将在重新讨论互相关和各态历经性后得出这个结论。另外各态历经性也是随机过程一个重要性质。

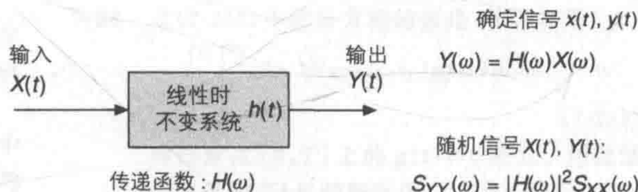


图 8-1 线性时不变系统中，确定信号和广义平稳随机信号的输入/输出频域关系

8.2 再论互相关

第 7 章研究了微分方程：

$$a_n\frac{\mathrm{d}^n}{\mathrm{d}t^n}X(t)+a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}X(t)+\cdots+a_1\frac{\mathrm{d}}{\mathrm{d}t}X(t)+a_0X(t)=Y(t) \tag{8-2}$$

其中$\{a_1,\cdots,a_n\}$是常系数，随机过程$Y(t)$是微分方程所表示系统的输入，$X(t)$是系统的输出(请注意在第 7 章和式(8-2)中，交换了$Y(t)$和$X(t)$)。因为$Y(t)$是随机的，所以一般而言不可能针对每个$X(t)$都写出一个表达式。我们首先研究它的均值：

$$a_n \frac{\mathrm{d}^n}{\mathrm{d}t^n}\mu_X(t)+a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}\mu_X(t)+\cdots+a_1\frac{\mathrm{d}}{\mathrm{d}t}\mu_X(t)+a_0\mu_X(t)=\mu_Y(t) \tag{8-3}$$

其中$t>0$，这里假设求导和数学期望可以交换顺序。解方程中的$\mu_X(t)$，可以得到输出$X(t)$平均表现的一个表达式。在很多应用中，对于所有t都有$\mu_Y(t)$为零，并且系统初始条件也是零，那么$\mu_X(t)$对于$t\in\mathcal{T}$也将恒为零。因此，对于$X(t)$的总体行为研究而言，均值的差分方程提供的信息用处很有限。

将$X(t)$中t用t_2代替，微分方程的每一项都乘以$X(t_1)$，并取数学期望，以研究自相关函数：

$$\begin{aligned}&a_n\frac{\partial^n}{\partial t_2^n}R_{XX}(t_1,t_2)+a_{n-1}\frac{\partial^{n-1}}{\partial t_2^{n-1}}R_{XX}(t_1,t_2)+\cdots+a_1\frac{\partial}{\partial t_2}R_{XX}(t_1,t_2)+a_0R_{XX}(t_1,t_2)\\&=R_{XY}(t_1,t_2)\end{aligned} \tag{8-4}$$

其中t_1，$t_2>0$。这表明可以通过求解以$R_{XY}(t_1,t_2)$为输入的微分方程获取$R_{XX}(t_1,t_2)$，而$R_{XY}(t_1,t_2)$可以通过求解下面的微分方程获取，这个微分方程通过式(8-2)乘以$Y(t_2)$并取数学期望得到：

$$a_2\frac{\partial^2}{\partial t_1^2}R_{XY}(t_1,t_2)+a_1\frac{\partial}{\partial t_1}R_{XY}(t_1,t_2)+a_0R_{XY}(t_1,t_2)=R_{YY}(t_1,t_2) \tag{8-5}$$

其中$t_1,t_2>0$。从式(8-5)中可以研究线性微分方描述的随机过程的二阶行为。特别地，如果这个随机过程具有零均值，它的方差函数可以通过分析自相关函数在$t_1=t_2=t$的数值得到。该结论也适用于线性差分方程所描述的随机序列。

广义平稳随机过程$X(t)$的自相关函数可以看成联合概率密度函数的数学期望：

$$\begin{aligned}R_{XX}(\tau)&\triangleq\mathcal{E}[X(t)X(t+\tau)]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2f_{X(t),X(t+\tau)}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\end{aligned} \tag{8-6}$$

考虑如下自相关函数的估计：

$$\overline{R}_{XX}(\tau)\triangleq\frac{1}{2T}\int_{-T}^{T}X(t)X(t+\tau)\mathrm{d}t \tag{8-7}$$

这里积分是均方意义下的。相应的均值的表达式为：

$$\overline{\mu}_X\triangleq\frac{1}{2T}\int_{-T}^{T}X(t)\mathrm{d}t \tag{8-8}$$

因为$X(t)$是随机过程，估计量$\overline{R}_{XX}(\tau)$和$\overline{\mu}_X$也是随机变量，所以可以计算它们的期望：

$$\mathcal{E}[\overline{\mu}_X]=\frac{1}{2T}\int_{-T}^{T}\mathcal{E}[X(t)]\mathrm{d}t=\mu_X \tag{8-9}$$

$$\mathcal{E}[\overline{R}_{XX}(\tau)]=\frac{1}{2T}\int_{-T}^{T}\mathcal{E}[X(t)X(t+\tau)]\mathrm{d}t=R_{XX}(\tau) \tag{8-10}$$

式(8-9)、式(8-10)说明这两个估计均*无偏*，即在均值意义上它们与实际数值相同(估计的性质请参见第 9 章)。

假设我们按时间平均计算某个现实$x(t)$，那么：

$$\overline{\overline{R}}_{XX}(\tau)=\frac{1}{2T}\int_{-T}^{T}x(t)x(t+\tau)\mathrm{d}t \tag{8-11}$$

$$\overline{\overline{\mu}}_X=\frac{1}{2T}\int_{-T}^{T}x(t)\mathrm{d}t \tag{8-12}$$

其中双横线符号表示这种类型的估计。由于$\overline{\overline{R}}_{XX}(\tau)$是$X(t)$的一个具体现实的平均，因此

是一个确定性自相关函数，$\overline{\mu}$也同样如此。比较式(8-11)中的确定性自相关函数和式(8-6)中的定义，发现它们有类似的形式。但是定义式(8-6)中的自相关函数通过联合概率密度函数对 $X(t)$在 t 和 $t+\tau$ 时刻的所有可能的数值加权，而式(8-11)中确定性自相关函数只在 $X(t)$的某一个现实上进行积分，无需加权。不同情形的自相关函数和均值分别如图 8-2 和图 8-3 所示。

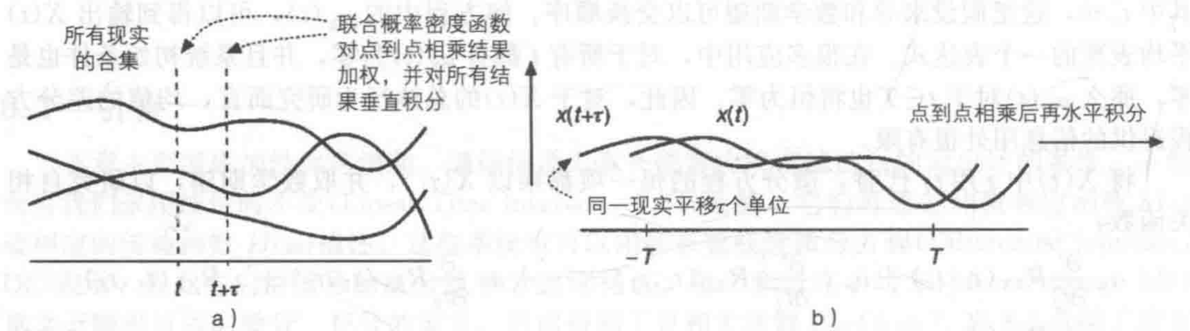


图 8-2 自相关函数。a）利用概率密度函数对 t 和 $t+\tau$ 两个时刻的所有可能结果加权后再积分，得到 $R_{XX}(\tau)$；b）对一个现实和其平移 τ 个单位的乘积在 $[-T,T]$ 上取平均，得到$\overline{R}_{XX}(\tau)$

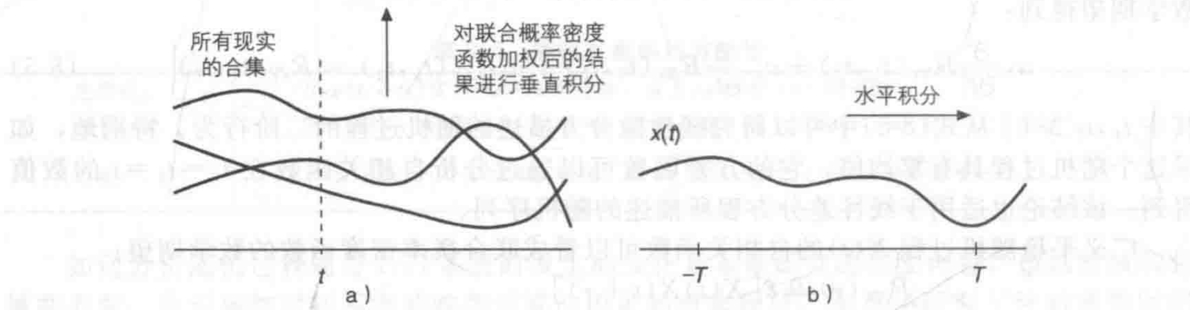


图 8-3 均值。a）利用概率密度函数对 t 时刻所有可能的结果进行加权，得到 μ_X；b）对一个现实在 $[-T,T]$ 上取平均，得到 $\overline{\mu}_X$

时间平均自相关函数是一种 $X(t)$的某一个现实和其时移之间相似性的测度。图 8-4 展示了一个假设的矩形现实和不同时移的时间平均自相关函数$\overline{R}_{XX}(\tau)$。它在 $\tau=0$ 时取得最大值，此时 $x(t)$和 $x(t+\tau)$最相似(完全一样)。当 τ 增加时$\overline{R}_{XX}(\tau)$平稳下降，意味着函数上两个波形越来越不相似。式(8-6)定义的统计自相关函数有类似的现象，但它是两个随机变量 $X(t)$和 $X(t+\tau)$的二阶矩的平均。区别是$\overline{R}_{XX}(\tau)$是时间上的平均，而 $R_{XX}(\tau)$是所有可能现实上的平均。第 7 章的主要目的是利用随机过程的概率描述替代寻找 $X(t)$的所有现实。

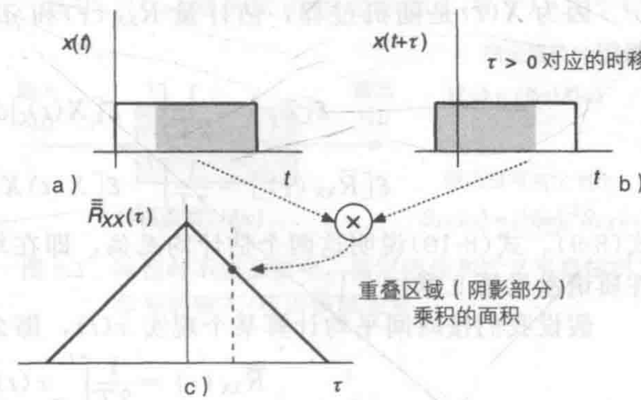


图 8-4 确定性自相关函数。a）随机过程 $X(t)$的一个现实 $x(t)$；b）该现实的时移；c）对应的时间平均自相关函数$\overline{R}_{XX}(\tau)$

因为随机过程是时间的函数，所以这种自相关函数引出的相似性测度隐含了随机过程在一定程度上的可预测性，使得过去的数值包含了一些关于未来数值的“信息”。大部分随机过程的信号处理算法利用互相关信息达到一定目的，比如在噪声污染的接收信号中估计发射信号。我们将针对随机信号和处理这种信号的系统进一步研究互相关函数。在

研究滤波器之前，先讨论各态历经性(ergodicity)这个概念，对于某些随机过程有：

$$\lim_{T\to\infty}\overline{\mu}_X=\mu_X,\quad \lim_{T\to\infty}\overline{R}_{XX}(\tau)=R_{XX}(\tau) \tag{8-13}$$

取极限后，$X(t)$的任一现实的时间平均与所有现实合集的期望相同。

8.3 各态历经性

首先考虑均值各态历经性。式(8-8)给出了随机过程 $X(t)$均值的时间平均 $\overline{\mu}_X$。如果 $X(t)$的均值为常数，则 $\varepsilon[\overline{\mu}_X]=\mu_X$，意味着即使是有限时间积分，$\overline{\mu}_X$ 仍然是 μ_X 的一个无偏估计。均值的各态历经性是非常重要的性质，意味着当 $T\to\infty$时，随机过程的时间平均收敛于均值。

定义(均值的各态历经性(随机过程))　设随机过程 $X(t)$均值为常数 μ_X，如果满足

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}X(t)\,\mathrm{d}t=\mu_X \tag{8-14}$$

则称 $X(t)$满足**均值各态历经性**。

时间平均趋于数学期望暗指时间平均的方差趋于零。这个特性很有用，也就是说，在足够长的时间对 $X(t)$的任一现实取平均，可以精确估计随机过程的均值。随机过程的均值各态历经性可以通过后面讨论的方法验证。

定理 8-1(均值的各态历经性)　设随机过程 $X(t)$的自协方差函数为 $C_{XX}(\tau)$，$X(t)$具有均值各态历经性当且仅当：

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)C_{XX}(\tau)\,\mathrm{d}\tau=0 \tag{8-15}$$

其中

$$C_{XX}(\tau)=R_{XX}(\tau)-\mu_X^2 \tag{8-16}$$

证明：考虑式(8-14)中积分式的方差：

$$\sigma_\mu^2=\mathcal{E}\left[\left(\frac{1}{2T}\int_{-T}^{T}X(t)\,\mathrm{d}t-\mu_X\right)^2\right]=\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}C_{XX}(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \tag{8-17}$$

因为积分在均方意义下存在，所以 $\sigma_\mu^2\to 0$。当时间平均的方差趋于零时，它的概率密度函数变为位于 μ_X 处的 δ 函数。因此只需要验证式(8-17)中的双重积分等价于式(8-15)中的一重积分。第 7 章讨论了包含自相关函数 $R_{XX}(\tau)$的均方积分与此类似的等价问题，相关细节不再赘述。图 7-16 中关于 $R_{XX}(\tau)$的结果也适用于 $C_{XX}(\tau)$，只是积分限从 $[t_0,t]$ 改为 $[-T,T]$。

因为定理提到的 $C_{XX}(\tau)$只依赖时间延迟 τ，所以 $X(t)$广义平稳：当 $\tau\triangleq t_2-t_1$ 时 $C_{XX}(t_1,t_2)=C_{XX}(t_2-t_1)=C_{XX}(\tau)$。$\tau\triangleq t_2-t_1$ 和 $\alpha\triangleq t_1+t_2$ 的变量替换对应于图 7-16 中的坐标旋转。τ 的积分限显然是 $[-2T,2T]$，而 α 的积分限依赖于 τ。从图 7-16 易知，为保证正方形的积分区域，α 的范围必是 $[-(2T-|\tau|),2T-|\tau|]$。相应的雅可比矩阵是(参考附录 G)：

$$\boldsymbol{J}=\begin{bmatrix}\partial\tau/\partial t_1 & \partial\tau/\partial t_2\\ \partial\alpha/\partial t_1 & \partial\alpha/\partial t_2\end{bmatrix}=\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix} \tag{8-18}$$

所以：

$$\begin{aligned}\sigma_\mu^2&=\frac{1}{4T^2|\boldsymbol{J}|}\int_{-2T}^{2T}\int_{-(2T-|\tau|)}^{2T-|\tau|}C_{XX}(\tau)\,\mathrm{d}\alpha\,\mathrm{d}\tau\\&=\frac{1}{8T^2}\int_{-2T}^{2T}2(2T-|\tau|)C_{XX}(\tau)\,\mathrm{d}\tau\\&=\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)C_{XX}(\tau)\,\mathrm{d}\tau\end{aligned} \tag{8-19}$$

这里 $J=\det(\boldsymbol{J})=2$ 是变换的雅可比行列式。当 $T\to\infty$时 $\sigma_\mu^2\to 0$，证毕。

虽然原始积分是二重积分，但由于自协方差函数是单变量函数，所以加权函数 $\tau=0$

的中心高度为1，宽度为$4T$。当坐标轴与τ方向对齐时，关于α的积分导致了这个三角形加权。这个加权保证了依变换后数轴积分和依原始数轴积分的结果相同，并充分利用图7-16上$C_{XX}(t_1,t_2)$在$t_1=t_2$的垂线上为常数的事实(请注意，关于$C_{XX}(t_1,t_2)$的积分限已经改为$[-T,T]$)。

由于式(8-15)不一定容易计算，因此再给出一个均值意义下各态历经性的充分条件。

定理8-2　设随机过程$X(t)$的自协方差函数为$C_{XX}(\tau)$，当

$$\int_{-\infty}^{\infty}|C_{XX}(\tau)|\mathrm{d}\tau<\infty \tag{8-20}$$

时，$X(t)$具有均值各态历经性。

证明：取式(8-19)的右边的绝对值：

$$\begin{aligned}\sigma_\mu^2&=\frac{1}{2T}\left|\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)C_{XX}(\tau)\mathrm{d}\tau\right|\\&\leqslant\frac{1}{2T}\int_{-2T}^{2T}\left|\left(1-\frac{|\tau|}{2T}\right)C_{XX}(\tau)\right|\mathrm{d}\tau\leqslant\frac{1}{2T}\int_{-2T}^{2T}|C_{XX}(\tau)|\mathrm{d}\tau\end{aligned} \tag{8-21}$$

如果式(8-20)成立，则因为分母上的$T\to\infty$，式(8-21)的右边趋于零。

定义(均值的各态历经性(序列))　设随机序列$X[k]$的均值为常数μ_X，如果满足：

$$\lim_{T\to\infty}\frac{1}{2T+1}\sum_{k=-T}^{T}X[k]=\mu_X \tag{8-22}$$

则称该序列具有**均值各态历经性**。

因为一共有$2T+1$个求和项，为了获得μ_X的无偏估计，所以除以$2T+1$。

自协方差函数$C_{XX}[m]$的相应条件为(参考习题8-1)：

$$\lim_{T\to\infty}\frac{1}{2T+1}\sum_{m=-2T}^{2T}\left(1-\frac{|m|}{2T+1}\right)C_{XX}[m]=0 \tag{8-23}$$

和

$$\sum_{m=-\infty}^{\infty}|C_{XX}[m]|<\infty \tag{8-24}$$

例8-1　设$X(t)$是不相关过程，当$\tau\neq0$时$R_{XX}(\tau)=\mathcal{E}[X(t)X(t+\tau)]=\mathcal{E}[X(t)]\mathcal{E}[X(t+\tau)]=\mu_X^2$，且方差$\sigma_X^2$为有限值，所以

$$\begin{aligned}C_{XX}(\tau)&=\begin{cases}\mathcal{E}[X^2(t)]-\mu_X^2, & \tau=0\\0, & \text{其他}\end{cases}\\&=\sigma_X^2\delta(\tau)\end{aligned} \tag{8-25}$$

观察：

$$\sigma_\mu^2=\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)\sigma_X^2\delta(\tau)\mathrm{d}\tau=(1/2T)\sigma_X^2 \tag{8-26}$$

这里使用了δ函数的筛选特性。当$T\to\infty$时这个结果显然趋于零，满足均值各态历经的充分条件

$$\int_{-\infty}^{\infty}\sigma_X^2\delta(\tau)\mathrm{d}\tau=\sigma_X^2<\infty \tag{8-27}$$ ◀

例8-2　考虑一个零均值随机过程，如果其自相关函数为$R_{XX}(\tau)=\exp(-|\tau|)=C_{XX}(\tau)$。则它满足均值各态历经性的充分条件

$$\int_{-\infty}^{\infty}|\exp(-|\tau|)|\mathrm{d}\tau=2\int_0^{\infty}\exp(-\tau)\mathrm{d}\tau=2<\infty \tag{8-28}$$

这个例子说明一般而言，充分条件更容易检验，如果充分条件不满足，就需检验式(8-15)的充要条件。出于完整性，有

$$\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)\exp(-|\tau|)\mathrm{d}\tau=\frac{1}{T}\int_0^{2T}[\exp(-\tau)-(\tau/2T)\exp(-\tau)]\mathrm{d}\tau$$

$$=-(1/T)\exp(-\tau)\big|_0^{2T}+(1/2T^2)(\tau+1)\exp(-\tau)\big|_0^{2T}$$
$$=(1/T)[1-\exp(-2T)]+(1/2T^2)[(2T+1)\exp(-2T)-1] \tag{8-29}$$

当 $T\to\infty$ 时该式趋于零。◀

各态历经性可以推广到其他的数学期望，我们将研究相关函数的各态历经性。这种从随机过程的单个现实中估计均值和自相关函数的方法在很多应用中都非常重要，特别是后面将要研究的线性系统。

定义（相关函数的各态历经性） 设随机过程 $X(t)$ 的自相关函数为 $R_{XX}(\tau)$，当满足以下条件时

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}X(t)X(t+\tau)\mathrm{d}t=R_{XX}(\tau) \tag{8-30}$$

称 $X(t)$ 具有**相关函数的各态历经性**。

即使有限时长条件下，某一个现实上的时间平均也是 $R_{XX}(\tau)$ 的无偏估计。

$$\mathcal{E}\left[\frac{1}{2T}\int_{-T}^{T}X(t)X(t+\tau)\mathrm{d}t\right]=R_{XX}(\tau) \tag{8-31}$$

与均值各态历经性的分析类似，我们将分析自相关函数的方差，并指出当 $T\to\infty$ 时，方差趋于零。为简化符号，记：

$$Y_\tau(t)\triangleq X(t)X(t+\tau) \tag{8-32}$$

其均值 $\mu_{Y_\tau}=R_{XX}(\tau)$，自协方差函数为：

$$C_{Y_\tau Y_\tau}(\alpha)=R_{Y_\tau Y_\tau}(\alpha)-R_{XX}^2(\tau) \tag{8-33}$$

对每一个 τ 单独定义一个 $Y_\tau(t)$，$Y_\tau(t)$ 的均值与 t 无关，而且由于它的自相关函数仅依赖单个变量 α，$Y_\tau(t)$ 广义平稳。

定理 8-3（相关函数的各态历经性） 设随机过程 $X(t)$ 具有相关函数的各态历经性，当且仅当：

$$\lim_{T\to\infty}\int_{-T}^{T}\left(1-\frac{|\alpha|}{2T}\right)C_{Y_\tau Y_\tau}(\alpha)\mathrm{d}\alpha=0 \tag{8-34}$$

证明：$Y_\tau(t)$ 的方差为

$$\sigma_{Y_\tau}^2=\mathcal{E}\left[\left|\frac{1}{2T}\int_{-T}^{T}Y_\tau(t)\mathrm{d}t-\mu_{Y_\tau}\right|^2\right] \tag{8-35}$$

和定理 8-1 类似，可以马上写出：

$$\sigma_{Y_\tau}^2=\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}C_{Y_\tau Y_\tau}(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2 \tag{8-36}$$

这里 $C_{Y_\tau Y_\tau}(t_1,t_2)\triangleq\mathcal{E}[Y_\tau(t_1)Y_\tau(t_2)]-\mu_{Y_\tau}^2=R_{Y_\tau Y_\tau}(t_1,t_2)-R_{XX}^2(\tau)$，经坐标变换可得到：

$$\sigma_{Y_\tau}^2=\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\alpha|}{2T}\right)C_{Y_\tau Y_\tau}(\alpha)\mathrm{d}\alpha \tag{8-37}$$

当 $T\to\infty$ 时，$\sigma_{Y_\tau}^2\to0$，证毕。

由于式(8-37)与均值各态历经性的要求类似，只是 $C_{XX}(\tau)$ 换成了 $C_{Y_\tau Y_\tau}(\alpha)$，所以式(8-38)是相关函数各态历经性的一个充分条件：

$$\int_{-\infty}^{\infty}C_{Y_\tau Y_\tau}(\alpha)\mathrm{d}\alpha<\infty \tag{8-38}$$

请注意，式(8-33)中 $R_{Y_\tau Y_\tau}(\alpha)$ 是原始随机过程的四阶矩：

$$\begin{aligned}R_{Y_\tau Y_\tau}(\alpha)&=\mathcal{E}[Y_\tau(t)Y_\tau(t+\alpha)]\\&=\mathcal{E}[X(t)X(t+\tau)X(t+\alpha)X(t+\tau+\alpha)]\end{aligned} \tag{8-39}$$

例 8-3 很难写出四阶矩 $R_{Y_\tau Y_\tau}(\alpha)$ 的表达式，但零均值高斯随机过程 $X(t)$ 的所有偶数阶矩均可用二阶矩表示。特别地：如果对于 $n=1,2,3,4$，$\{X(t_n)\}$ 是联合高斯随机过程，则有：

$$\begin{aligned}\mathcal{E}[X(t_1)X(t_2)X(t_3)X(t_4)] = & \mathcal{E}[X(t_1)X(t_2)]\mathcal{E}[X(t_3)X(t_4)] \\ & + \mathcal{E}[X(t_1)X(t_3)]\mathcal{E}[X(t_2)X(t_4)] \\ & + \mathcal{E}[X(t_1)X(t_4)]\mathcal{E}[X(t_2)X(t_3)]\end{aligned} \tag{8-40}$$

将该结果应用于 $R_{Y_\tau Y_\tau}(\alpha)$，有

$$\begin{aligned}R_{Y_\tau Y_\tau}(\alpha) &= \mathcal{E}[X(t)X(t+\tau)X(t+\alpha)X(t+\tau+\alpha)] \\ &= R_{XX}^2(\tau) + R_{XX}^2(\alpha) + R_{XX}(\tau+\alpha)R_{XX}(\alpha-\tau)\end{aligned} \tag{8-41}$$

和

$$C_{Y_\tau Y_\tau}(\alpha) = R_{XX}^2(\alpha) + R_{XX}(\tau+\alpha)R_{XX}(\alpha-\tau) \tag{8-42}$$

习题 8-6 将研究相关函数为例 8-2 中的 $R_{XX}(\tau)$ 的零均值高斯随机过程 $X(t)$ 是否具有相关函数各态历经性。 ◀

与研究均值和相关函数各态历经性类似，为了用时间平均代替数学期望，我们可以定义分布的各态历经性。第 5 章指出，随机变量的累积分布函数可以用指示函数的期望表示。将该方法推广到随机过程：

$$\mathcal{E}[I_{(-\infty,x]}(X(t))] = F_{X(t)}(x) \tag{8-43}$$

由此，我们给出如下定义：

定义(分布的各态历经性) 对于任意 x，随机过程 $X(t)$ 具有**分布的各态历经性**，当且仅当：

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} I_{(-\infty,x]}(X(t))\mathrm{d}t = F_{X(t)}(x) \tag{8-44}$$

这个定义阐述了通过对指示函数的不同现实取平均，可以估计分布。假设指示函数的变量$X(t)$是独立同分布的标准高斯随机过程。$X(t)$的一个现实例子见图 8-5a。如果式(8-43)中 $x=\infty$，相应 $I_{(-\infty,\infty)}(X(t))$的现实显然是常数 1，类似地，如果 $x=-\infty$，现实恒为零。对于所有其他 x，$I_{(-\infty,x]}(X(t))$的现实将在 0 和 1 之间变化。这些现实中会有很多跃变，也有一些相对平坦的区域。比如 $x=0$ 时，$I_{[0,\infty)}(X(t))$的 0 和 1 取值的持续时间基本相当，图 8-5b给出了一个这样的现实。当我们取这个现实的平均值时，值为 0.5，也就是 $F_{X(t)}(0)$。对于$x>0$，时间平均应该>0.5，对于 $x<0$，时间平均应该<0.5。图 8-5c 中给出了 $x=2$ 时 $I_{(-\infty,2]}(X(t))$的一个现实。指示函数记录了何时 $X(t)$大于 x，由此可以将累积分布函数作为二元过程的期望。

利用前面均值和相关函数各态历经性的结果，我们可以得到分布各态历经性的充要条件。为简化符号，定义随机过程

$$Z_x(t) \triangleq I_{(-\infty,x]}(X(t)) \tag{8-45}$$

其中指示函数的下标 x 表示特定值。它的期望为：

$$\begin{aligned}\mathcal{E}[Z_x(t)] &= \int_{-\infty}^{\infty} I_{(-\infty,x]}(x(t))f_{X(t)}(x(t))\mathrm{d}x(t) \\ &= \int_{-\infty}^{x} f_{X(t)}(x(t))\mathrm{d}x(t) = F_{X(t)}(x)\end{aligned} \tag{8-46}$$

这是累积分布函数的无偏估计。自相关函数为：

$$\begin{aligned}R_{Z_xZ_x}(\tau) &= \mathcal{E}[I_{(-\infty,x]}(X(t))I_{(-\infty,x]}(X(t+\tau))] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{(-\infty,x]}(x(t))I_{(-\infty,x]}(x(t+\tau)) \\ &\quad \cdot f_{X(t),X(t+\tau)}(x(t),x(t+\tau))\mathrm{d}x(t)\mathrm{d}x(t+\tau) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{x} f_{X(t),X(t+\tau)}(x(t),x(t+\tau))\mathrm{d}x(t)\mathrm{d}x(t+\tau) \\ &= P(X(t)\leqslant x, X(t+\tau)\leqslant x) = F_{X(t),X(t+\tau)}(x,x)\end{aligned} \tag{8-47}$$

这是变量 x 相同条件下，随机过程 $X(t)$两个不同时刻的联合累积分布函数。由于

$X(t)$广义平稳，所以 $Z_x(t)$也类似地广义平稳，使得 $R_{Z_xZ_x}(\tau)$只依赖于时间延迟 τ。相应的自协方差函数为：

$$C_{Z_xZ_x}(\tau) = R_{Z_xZ_x}(\tau) - F_{X(t)}^2(x) \tag{8-48}$$

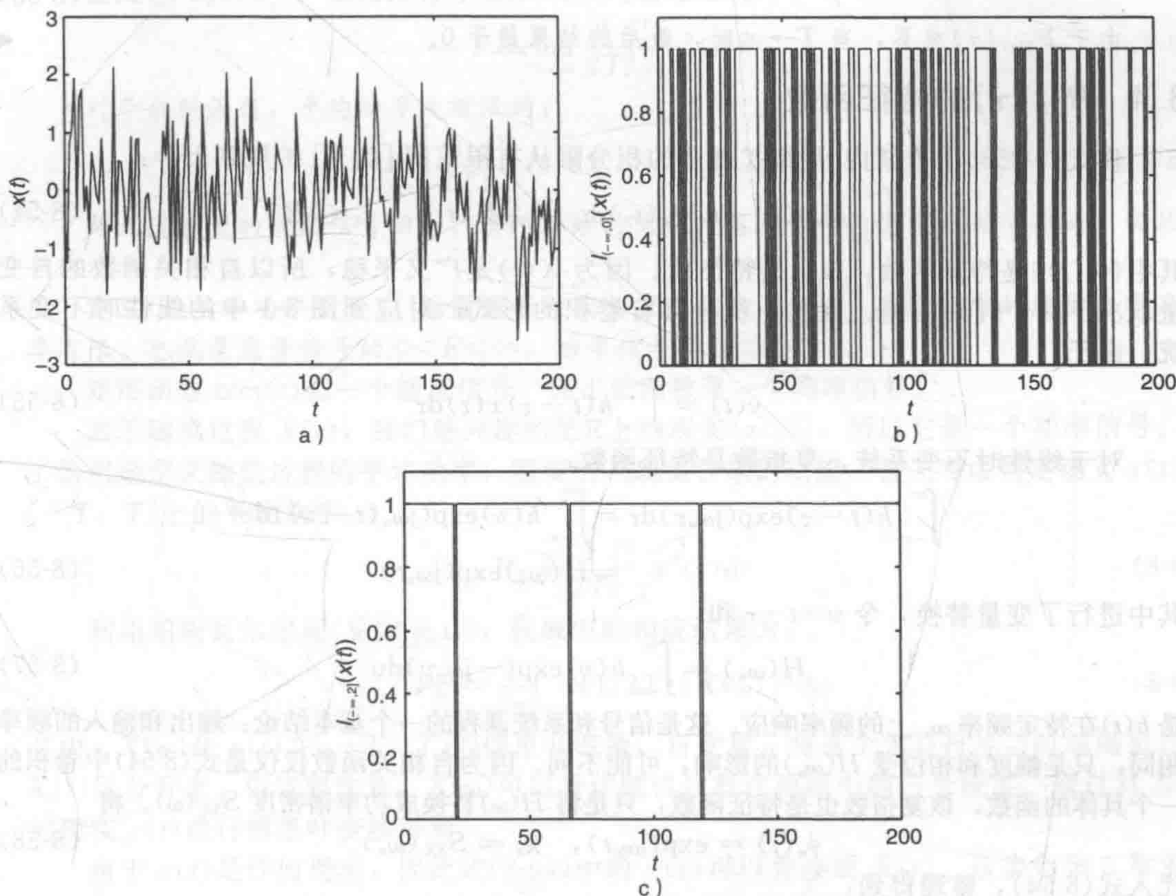


图 8-5 随机过程和指示函数。a) 高斯随机过程的一个现实；b) $I_{(-\infty,0]}(X(t))$的现实；c) $I_{(-\infty,2]}(X(t))$的现实

定理 8-4(分布的各态历经性) 对于任意 x，随机过程 $X(t)$具有分布的各态历经性，当且仅当：

$$\lim_{T\to\infty}\int_{-2T}^{2T}\left(1-\frac{\tau}{2T}\right)C_{Z_xZ_x}(\tau)\mathrm{d}\tau = 0 \tag{8-49}$$

证明：除了用 $C_{Z_xZ_x}(\tau)$替代 $C_{XX}(\tau)$外，证明过程与均值各态历经性的证明类似，证毕。

分布各态历经性的一个充分条件是：

$$\int_{-\infty}^{\infty}|C_{Z_xZ_x}(\tau)|\mathrm{d}\tau < \infty \tag{8-50}$$

例 8-4 若随机过程 $X(t)$独立同分布，则该过程具有分布各态历经性(也具有均值各态历经性)。由于自相关函数 $R_{Z_xZ_x}(\tau)=F_{X(t)}(x)F_{X(t+\tau)}(x)$，因此对于任意 $\tau\neq0$，由于 $F_{X(t+\tau)}(x)=F_{X(t)}(x)$，故 $C_{Z_xZ_x}(\tau)=0$。当 $\tau=0$ 时，

$$\begin{aligned}R_{Z_xZ_x}(0) &= \mathcal{E}[I_{(-\infty,x]}^2(X(t))] = \mathcal{E}[I_{(-\infty,x]}(X(t))]\\ &= \int_{-\infty}^{x} f_{X(t)}(x(t))\mathrm{d}x(t) = P(X(t)\leqslant x) = F_{X(t)}(x)\end{aligned} \tag{8-51}$$

$$C_{Z_xZ_x}(\tau) = [F_{X(t)}(x) - F_{X(t)}^2(x)]\delta(\tau) = F_{X(t)}(x)[1-F_{X(t)}(x)]\delta(\tau) \tag{8-52}$$

$C_{Z_xZ_x}(0)$是 $Z_x(t)$的方差，在这道例题中对应 $P(X(t)\leqslant x)P(X(t)>x)$。因此，$t$ 固定后，

$Z_2(t)$可以看成成功概率为$F_{X(t)}(x)$的伯努利随机变量。代入式(8-49)中，得到

$$\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)F_{X(t)}(x)[1-F_{X(t)}(x)]\delta(\tau)\mathrm{d}\tau=(1/2T)F_{X(t)}(x)[1-F_{X(t)}(x)] \tag{8-53}$$

由于$F_{X(t)}(x)$有界，当$T\to\infty$时，最后的结果趋于0。◀

8.4　$R_{XX}(\tau)$的特征函数

将式(7-258)中卡胡内-列维扩展式的积分限从有限范围$[0,T]$扩展到$\pm\infty$：

$$\int_{-\infty}^{\infty}R_{XX}(t_1-t_2)\phi_n(t_2)\mathrm{d}t_2=\lambda_n\phi_n(t_1),\quad t_1\in R \tag{8-54}$$

其中$\{\phi_n(t)\}$是特征函数，$\{\lambda_n\}$是特征值。因为$X(t)$是广义平稳，所以自相关函数的自变量取决于两个时刻的差。因此，积分具有卷积的形式，对应到图8-1中的线性时不变系统，由于

$$y(t)=\int_{-\infty}^{\infty}h(t-\tau)x(\tau)\mathrm{d}\tau \tag{8-55}$$

对于线性时不变系统，复指数是特征函数：

$$\int_{-\infty}^{\infty}h(t-\tau)\exp(\mathrm{j}\omega_n\tau)\mathrm{d}\tau=\int_{-\infty}^{\infty}h(v)\exp(\mathrm{j}\omega_n(t-v))\mathrm{d}v=H(\omega_n)\exp(\mathrm{j}\omega_n t) \tag{8-56}$$

其中进行了变量替换，令$v=t-\tau$和

$$H(\omega_n)=\int_{-\infty}^{\infty}h(v)\exp(-\mathrm{j}\omega_n v)\mathrm{d}v \tag{8-57}$$

是$h(t)$在特定频率ω_n上的频率响应。这是信号和系统课程的一个基本结论：输出和输入的频率相同，只是幅度和相位受$H(\omega_n)$的影响，可能不同。因为自相关函数仅仅是式(8-54)中卷积的一个具体的函数，该复指数也是特征函数，只是将$H(\omega)$替换成功率谱密度$S_{XX}(\omega)$。将

$$\phi_n(t)=\exp(\mathrm{j}\omega_n t),\quad \lambda_n=S_{XX}(\omega_n) \tag{8-58}$$

代入式(8-54)，整理得到：

$$\int_{-\infty}^{\infty}R_{XX}(t_1-t_2)\exp(-\mathrm{j}\omega_n(t_1-t_2))\mathrm{d}t_2=S_{XX}(\omega_n) \tag{8-59}$$

令$\tau\triangleq t_1-t_2$且$\omega_n\to\omega$，我们发现$S_{XX}(\omega)$是$R_{XX}(\tau)$的傅里叶变换，正如在引言中提到的：

$$\int_{-\infty}^{\infty}R_{XX}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau=S_{XX}(\omega) \tag{8-60}$$

与$X(t)$的卡胡内-列维扩展式中在$\mathcal{T}=[0,T]$积分获得可数个特征函数不同，在$\mathcal{T}=\mathcal{R}$上得到的特征函数的数量不可数。由第7章可知，如果$R_{XX}(\tau)$的周期为T，则该过程可以做傅里叶级数展开。本节中的结论可以认为是将傅里叶级数的结果扩展到非周期函数$R_{XX}(\tau)$。然而，值得注意的是，虽然式(8-54)是一个卷积，但是$R_{XX}(\tau)$没有用作$X(t)$的滤波器。在8.5节中，我们将通过不同的方法推导功率谱密度，并说明它与随机过程的功率之间的关系。功率谱密度有很多性质，前面图8-1给出了当$X(t)$由$H(\omega)$滤波后，$S_{YY}(\omega)$和$S_{XX}(\omega)$的关系，本章稍后将进行证明。

8.5　功率谱密度

简要回顾非随机信号的功率和能量。

定义(能量)　确定性信号$x(t)$的**能量**定义为：

$$E\triangleq\int_{-\infty}^{\infty}x^2(t)\mathrm{d}t \tag{8-61}$$

如果$x(t)$在$\mathcal{R}$上是周期信号，则E无穷大。例如，对于$x(t)=\cos(\omega t)$：

$$E = \int_{-\infty}^{\infty} \cos^2(\omega t)\mathrm{d}t = \int_{-\infty}^{\infty} (1/2)[1 + \cos(2\omega t)]\mathrm{d}t = \infty$$

对于这样的信号，考虑用平均功率。

定义(平均功率) 确定性信号 $x(t)$ 的**平均功率**定义为：

$$P \triangleq \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x^2(t)\mathrm{d}t \tag{8-62}$$

对于余弦函数，平均功率是有限的：

$$P = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} (1/2)[1 + \cos(2\omega t)]\mathrm{d}t = 1/2 \tag{8-63}$$

与无限长度的周期信号不同，有限持续时间的非随机信号的能量 E 是有限的，因此其平均功率 P 为 0。

定义(能量和功率信号) **能量信号**的能量有限且功率为 0。**功率信号**的能量无限且功率有限。也就是能量信号的 $0<E<\infty$，功率信号的 $0<P<\infty$。

矩形函数 $\mathrm{rect}(t)$ 是一个能量信号，而正弦函数是一个功率信号。

对于随机过程 $X(t)$，我们感兴趣的是 $\mathcal{R}$ 上的现实 $\{x(t)\}$，所以它是一个功率信号。为了恰当地定义随机过程的平均功率，需要引入现实合集的期望。首先考虑特定现实 $x(t)$ 在 $[-T, T]$ 上的平均功率：

$$P_T \triangleq \frac{1}{2T}\int_{-T}^{T} x^2(t)\mathrm{d}t \tag{8-64}$$

利用帕斯瓦尔定理(见附录 C)，频域中的相应结果为：

$$P_T = \frac{1}{2\pi}\int_{-\infty}^{\infty} (1/2T)\,|X(\omega)|^2\mathrm{d}\omega \tag{8-65}$$

其中，$X(\omega)$ 是 $x(t)$ 在 $[-T, T]$ 的傅里叶变换。自变量 ω 将这个大写符号与时域随机过程 $X(t)$ 区分开来。但需要注意，$X(\omega)$ 可以看作随机过程在频域中的一个现实。通过对 $X(t)$ 的现实 $x(t)$ 进行傅里叶变换得到。

由于 $x(t)$ 是任何现实，因此式(8-64)中的 $x(t)$ 可以替换成 $X(t)$。这类似第 5 章求条件期望时，条件从具体值变成随机变量。我们最初研究了期望 $\mathcal{E}[X|Y=y]$，它是随机变量取值 y 的函数。将条件扩展成随机变量：$\mathcal{E}[X|Y]$ 可以看成是随机变量 Y 的函数。类似地，式(8-64)是一个现实的函数，我们希望将其推广为随机过程 $X(t)$ 的函数。因此，式(8-65)中现实 $X(\omega)$ 可以替换成相应随机过程的傅里叶变换。式(8-64)、式(8-65)中的 P_T 变成了随机变量。

将随机过程的平均功率定义成 P_T 的期望，并令 $T\to\infty$：

$$P_{XX} \triangleq \lim_{T\to\infty} \mathcal{E}[P_T] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \lim_{T\to\infty}(1/2T)\,\mathcal{E}[\,|X(\omega)|^2]\mathrm{d}\omega \tag{8-66}$$

其中，假定极限和积分顺序可以互换。定义极限：

$$S_{XX}(\omega) \triangleq \lim_{T\to\infty}(1/2T)\,\mathcal{E}[\,|X(\omega)|^2] \tag{8-67}$$

从而

$$P_{XX} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XX}(\omega)\mathrm{d}\omega \tag{8-68}$$

最后一个表达式说明 $S_{XX}(\omega)$ 是功率谱密度：当对它在 ω 的一个区间积分时，可以得到这些频率的平均功率。随机过程在整个区间 $\omega\in\mathcal{R}$ 的平均功率是 P_{XX}。功率谱密度 $S_{XX}(\omega)$ 与 8.4 节讨论 $R_{XX}(\tau)$ 的特征函数时用到的功率谱密度一样，它的单位是瓦特/弧度。

最后一步是推导 $S_{XX}(\omega)$ 的有效表达式。将 $|X(\omega)|^2 = X^*(\omega)X(\omega)$ 中每一项的傅里叶变换代入式(8-67)，得到(上标 * 表示复共轭)：

$$S_{XX}(\omega)=\lim_{T\to\infty}\frac{1}{2T}\mathcal{E}\left[\int_{-T}^{T}X(t_1)\exp(-\mathrm{j}\omega t_1)\mathrm{d}t_1\int_{-T}^{T}X(t_2)\exp(\mathrm{j}\omega t_2)\mathrm{d}t_2\right]$$
$$=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T}R_{XX}(t_1,t_2)\exp(-\mathrm{j}\omega(t_1-t_2))\mathrm{d}t_1\mathrm{d}t_2 \tag{8-69}$$

其中，使用了不同的积分变量并假设随机过程 $X(t)$ 为实数。取极限前，将 $t_1,t_2\in[-T,T]$ 上的自相关函数定义成 $R_{XX}(t_1,t_2)\triangleq\mathcal{E}[X(t_1)X(t_2)]$。并假设 $X(t)$ 广义平稳，从而 $R_{XX}(t_1,t_2)=R_{XX}(t_2-t_1)$ 并令 $\tau\triangleq t_1-t_2$。与各态历经性中的变量变换相同，功率谱密度变成：

$$S_{XX}(\omega)=\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T}(2T-|\tau|)R_{XX}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau$$
$$=\lim_{T\to\infty}\int_{-2T}^{2T}R_{XX}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau$$
$$-\lim_{T\to\infty}\int_{-2T}^{2T}(1/2T)|\tau|R_{XX}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau \tag{8-70}$$

如果满足下列有界条件：

$$\int_{-\infty}^{\infty}|\tau|R_{XX}(\tau)\mathrm{d}\tau<\infty \tag{8-71}$$

上面第二项积分的极限为 0，得到

$$S_{XX}(\omega)=\int_{-\infty}^{\infty}R_{XX}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau \tag{8-72}$$

这与式(8-60)利用特征函数得到的结果相同。式(8-71)中的条件要求自相关函数迅速衰减到 0，这意味着该随机过程不能“太相关。”我们将这些结论归纳为下面的定义。

定义(功率谱密度)　自相关函数为 $R_{XX}(\tau)$ 的广义平稳过程 $X(t)$ 的**功率谱密度**(Power Spectral Density，PSD)为

$$S_{XX}(\omega)=\int_{-\infty}^{\infty}R_{XX}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau \tag{8-73}$$

其中 ω 是角频率，$\omega=2\pi f$。从频率的角度，功率谱密度等于

$$S_{XX}(f)=\int_{-\infty}^{\infty}R_{XX}(\tau)\exp(-\mathrm{j}2\pi f\tau)\mathrm{d}\tau \tag{8-74}$$

自相关函数是功率谱密度的傅里叶反变换：

$$R_{XX}(\tau)=\frac{1}{2\pi}\int_{-\pi}^{\pi}S_{XX}(\omega)\exp(\mathrm{j}\omega\tau)\mathrm{d}\omega \tag{8-75}$$

代入 $\omega=2\pi f$，得到：

$$R_{XX}(\tau)=\int_{-\infty}^{\infty}S_{XX}(f)\exp(\mathrm{j}2\pi f\tau)\mathrm{d}f \tag{8-76}$$

为了完整性，我们给出其他随机信号和变换类型对应的其他形式的功率谱密度。

- 随机过程，拉普拉斯变换：

$$S_{XX}(s)=\int_{-\infty}^{\infty}R_{XX}(\tau)\exp(-s\tau)\mathrm{d}\tau \tag{8-77}$$
$$R_{XX}(\tau)=\frac{1}{2\pi\mathrm{j}}\int_{c-\mathrm{j}\infty}^{c+\mathrm{j}\infty}S_{XX}(s)\exp(s\tau)\mathrm{d}s \tag{8-78}$$

- 随机序列，离散时间傅里叶变换(DTFT)：

$$S_{XX}(\mathrm{j}\omega)=\sum_{m=-\infty}^{\infty}R_{XX}[m]\exp(-\mathrm{j}\omega m) \tag{8-79}$$
$$R_{XX}[m]=\frac{1}{2\pi}\int_{-\pi}^{\pi}S_{XX}(\mathrm{j}\omega)\exp(\mathrm{j}\omega m)\mathrm{d}\omega \tag{8-80}$$

- 随机序列，z 变换：

$$S_{XX}(z)=\sum_{m=-\infty}^{\infty}R_{XX}[m]z^{-m} \tag{8-81}$$

$$R_{XX}[m]=\frac{1}{2\pi \mathrm{j}}\oint_C S_{XX}(z)z^{m-1}\mathrm{d}z \tag{8-82}$$

表 8-2 给出了一些自相关函数以及它们的功率谱密度的例子。附录 G 讨论了拉普拉斯和 z 变换的收敛域(Regions Of Convergence，ROC)。图 8-6 给出了一个指数自相关函数和其功率谱密度的例子。当我们对随机过程的频率特性感兴趣时，一般研究 ω 或 f 形式的功率谱密度。而当研究输入为随机信号的系统时，一般使用拉普拉斯或 z 变换形式的功率谱密度。

表 8-2　功率谱密度的例子

$R_{XX}(\tau)$	$S_{XX}(\omega)$	$S_{XX}(s)$
白噪声：$\delta(\tau)$	1	1
指数：$\exp(-\alpha\lvert\tau\rvert)$	$2\alpha/(\alpha^2+\omega^2)$	$2\alpha/(\alpha^2-s^2)$
矩形：$\mathrm{rect}(\alpha\tau)$	$(1/\lvert\alpha\rvert)\mathrm{sinc}(\omega/2\pi\alpha)$	$(2\alpha/\lvert\alpha\rvert)\sinh(s/2\alpha)/s$
三角：$1-\alpha\lvert\tau\rvert/T$	$(1/\lvert\alpha\rvert)\mathrm{sinc}^2(\omega/2\pi\alpha)$	$(4\alpha^2/\lvert\alpha\rvert)\sinh^2(s/2\alpha)/s^2$
$R_{XX}[m]$	$S_{XX}(\mathrm{j}\omega)$	$S_{XX}(z)$
白噪声：$\delta[m]$	1	1
指数：$\alpha^{\lvert m\rvert}$	$(1-\alpha^2)/[1+\alpha^2-2\alpha\cos(\omega)]$	$(1-\alpha^2)/[(1-\alpha z^{-1})(1-\alpha z)]$

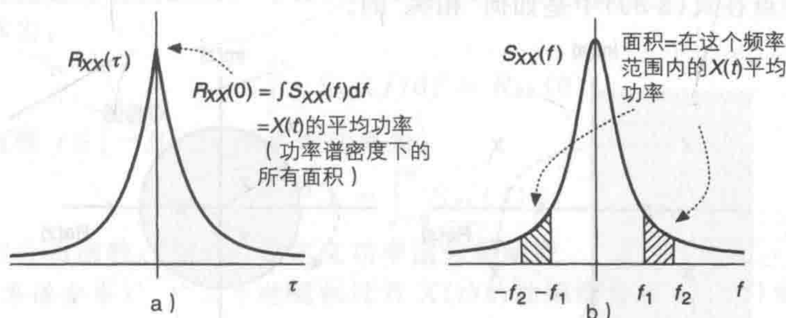


图 8-6　自相关函数和功率谱密度。a) 指数 $R_{XX}(\tau)$；b) $S_{XX}(f)$(W/Hz)，特定频率范围内的平均功率，通过对该范围内的功率谱密度积分得到

功率谱密度的性质如下：

- 平均功率。$X(t)$的二阶矩为

$$\mathcal{E}[X^2(t)]=R_{XX}(0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_{XX}(\omega)\mathrm{d}\omega \tag{8-83}$$

令式(8-76)的 $\tau=0$ 时得到上式。

- 实数。$S_{XX}(\omega)$是关于 ω 的实函数。由于 $R_{XX}(\tau)$是 τ 的偶函数，可得：

$$S_{XX}(\omega)=\int_0^{\infty}R_{XX}(\tau)[\exp(-\mathrm{j}\omega\tau)+\exp(\mathrm{j}\omega\tau)]\mathrm{d}\tau$$

$$=2\int_0^{\infty}R_{XX}(\tau)\cos(\omega\tau)\mathrm{d}\tau \tag{8-84}$$

即与 j 无关。

- 偶函数。由于式(8-84)中的 $\cos(\omega\tau)$是偶函数，因此 $S_{XX}(\omega)$是偶函数。
- 非负。从式(8-67)的原始定义可以得出 $S_{XX}(\omega)\geqslant 0$。
- 白噪声。如果 $X(t)$是一个白噪声过程，那么其自相关函数 $R_{XX}(\tau)=\delta(\tau)$，则$S_{XX}(\omega)=1$。

利用 δ 函数的筛选特性可以得出该结论：

$$S_{XX}(\omega)=\int_{-\infty}^{\infty}\delta(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau=1 \tag{8-85}$$

本章稍后将讨论白噪声。

- 推导过程。如果 $Y(t)=X'(t)$，则：

$$S_{YY}(\omega) = \omega^2 S_{XX}(\omega) \tag{8-86}$$

第 7 章中得出 $R_{YY}(\tau)=-\mathrm{d}^2 R_{XX}(\tau)/\mathrm{d}\tau^2$，再利用傅里叶变换的微分性质，可以得出该结论。

- 谱分解定理。如果 $S_{XX}(s)$是 s 的有理函数，它可以分解如下：

$$S_{XX}(s) = S_{XX}^{+}(s)S_{XX}^{-}(s) \tag{8-87}$$

其中 $S_{XX}^{+}(s)$的所有极零点在 s 平面的左半平面，$S_{XX}^{-}(s)$的所有极零点在 s 平面的右半平面。8.6 节将说明 $S_{XX}(s)$可以认为是功率谱密度平坦的输入信号通过传递函数为 $H(s)$的因果滤波器的输出。它是乘积的形式，

$$S_{XX}(s) = H(s)H(-s) \tag{8-88}$$

因为 $H(s)$稳定，所以它和 $S_{XX}^{+}(s)$的所有极点都在 s 平面的左半平面。同样地，反因果系统对应的 $H(-s)$和 $S_{XX}^{-}(s)$的所有极点都在 s 平面的右半平面。(假定 $H(s)$为绝对稳定，无极点在虚轴上。)

- 极点对称性。由谱分解性质，$S_{XX}(s)$的极点具有下列对称性：

$$\text{极点在 } s=-\sigma_1+\mathrm{j}\omega_1 \Rightarrow \text{极点在 } s=\sigma_1-\mathrm{j}\omega_1 \tag{8-89}$$

从图 8-7a 中可以直观看出这种双对称性质。关于实轴对称是由于自相关函数是偶含数，极点需要共轭。关于虚轴对称是由于 $S_{XX}(s)$是 ω 的偶函数。图中虚线显示了由于分解性质，两个复极点在式(8-89)中是如何"相关"的。

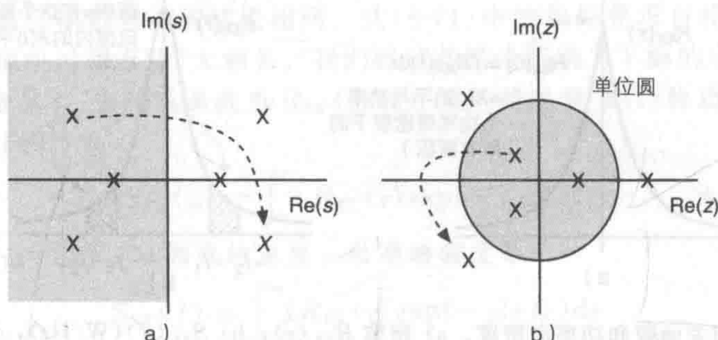


图 8-7 功率谱密度的极点位置。a) s 平面，$S_{XX}(s)=S_{XX}^{+}(s)S_{XX}^{-}(s)$的极点；b) z 平面，$S_{XX}(z)=S_{XX}^{+}(z)S_{XX}^{-}(z)$的极点，阴影区域中的极点对应功率谱密度的因果部分

- 零点对称性。当然，稳定性对 $H(s)$的零点没有限制。但是，$S_{XX}^{+}(s)$的所有零点在 s 平面的左半平面将方便稍后 11 章的学习，因此零点具有与极点相同的对称性。零点需要满足上述要求的原因是第 11 章介绍的维纳滤波要求 $S_{XX}^{+}(s)$可以取倒数，则零点将变成极点，为满足稳定性，零点需要在 s 平面的左半平面。(由于这个原因，我们也假设没有零点在虚轴上。)传递函数的所有零点在 s 平面的左半平面的系统称为最小相位系统。当 $S_{XX}^{+}(s)$的所有零点都不在 s 平面的右半平面时，群延迟较小。

随机序列的功率谱密度 $S_{XX}(\mathrm{j}\omega)$的对应性质留作练习。这里给出 $S_{XX}(z)$的谱分解性质：

$$S_{XX}(z) = S_{XX}^{+}(z)S_{XX}^{-}(z) \tag{8-90}$$

其中 $S_{XX}^{+}(z)$的极零点在 z 平面的单位圆内，$S_{XX}^{-}(z)$的极零点在单位圆外。传递函数 $H(z)$对应到式(8-88)的结果为：

$$S_{XX}(z) = H(z)H(z^{-1}) \tag{8-91}$$

极点对称性是

$$\text{极点在 } z=r_1\exp(\mathrm{j}\omega_1)=r_1[\cos(\omega_1)+\mathrm{j}\sin(\omega_1)]$$

$$\Rightarrow \text{极点在} z=(1/r_1)\exp(-\mathrm{j}\omega_1)=(1/r_1)(\cos(\omega_1)-\mathrm{j}\sin(\omega_1)) \tag{8-92}$$

这也适用于零点。图 8-7b 给出了一些例子，其中虚线表示由于分解性质，两个复极点在式(8-92)中是如何“相关”的。

虽然后面章节讨论了不同形式的功率谱密度，如 $S_{XX}(\omega)$、$S_{XX}(f)$、$S_{XX}(s)$、$S_{XX}(\mathrm{j}\omega)$和 $S_{XX}(z)$，我们通常只给出一种形式的结果，其他形式的结果可以同理得到。

8.6 功率谱分布

在第 3 章随机变量中，我们定义了随机变量 X 的累积分布函数 $F_X(x)$和概率密度函数 $f_X(x)$，它们的关系是：

$$f_X(x)=\frac{\mathrm{d}}{\mathrm{d}x}F_X(x) \tag{8-93}$$

假设 $F_X(x)$可导。在第 5 章中，定义了随机变量 X 的函数 $g(X)$的期望：

$$\mathcal{E}[g(X)]=\int_{-\infty}^{\infty} g(x)\mathrm{d}F_X(x) \tag{8-94}$$

这是一个积分变量为 $F_X(x)$的黎曼-斯蒂尔切斯积分。累积分布函数 $F_X(x)=P(X\leqslant x)$表示一个特定概率，而 $f_X(x)$是密度函数，对它积分可以得到概率：

$$P(X\leqslant x)=\int_{-\infty}^{x} f_X(x)\mathrm{d}x \tag{8-95}$$

功率谱密度也是密度函数，对它积分可以得到广义平稳过程 $X(t)$的平均功率。例如，整体平均功率为：

$$\int_{-\infty}^{\infty} S_{XX}(f)\mathrm{d}f=R_{XX}(0) \tag{8-96}$$

在一定频率范围 $f\in[-B, B]$的平均功率为：

$$P_{XX}=\int_{-B}^{B} S_{XX}(f)\mathrm{d}f \tag{8-97}$$

类似累积分布函数，我们可以定义功率谱分布函数。

定义（功率谱分布） 广义平稳随机过程 $X(t)$的**功率谱分布** $S_{XX}^{D}(f)$定义为：

$$R_{XX}(\tau)=\int_{-\infty}^{\infty} \exp(\mathrm{j}2\pi f\tau)\mathrm{d}S_{XX}^{D}(f) \tag{8-98}$$

其中 $R_{XX}(\tau)$是自相关函数。

为了与符号 $f_X(x)$和 $F_X(x)$一致，可以用 $s_X(f)$和 $S_X(f)$分别表示功率谱密度和功率谱分布函数。然而，大写 $S_{XX}(f)$在工程中通常表示功率谱密度，因此我们用上标表示功率谱分布 $S_{XX}^{D}(f)$。

注意，式(8-98)与 $\omega=2\pi f$ 的特征函数（见第 5 章）有相同的形式：

$$\Phi_X(\omega)=\int_{-\infty}^{\infty} \exp(\mathrm{j}\omega x)\mathrm{d}F_X(x) \tag{8-99}$$

虽然式(8-98)是一个傅里叶反变换，而 $\Phi_X(\omega)$是傅里叶变换（用 $-\omega$ 代替 ω）。由于 $S_{XX}^{D}(f)$是一个分布，它具有下列性质：

- 极限。

$$\lim_{f\to-\infty} S_{XX}^{D}(f)=0,\lim_{f\to\infty} S_{XX}^{D}(f)/R_{XX}(0)=1 \tag{8-100}$$

- 递增。对于 $f_1\leqslant f_2$：

$$S_{XX}^{D}(f_1)\leqslant S_{XX}^{D}(f_2) \tag{8-101}$$

- 右连续。

唯一需要注意的是，因为随机过程的总平均功率不限定为 1，$S_{XX}^{D}(f)$必须经过 $R_{XX}(0)=\int_{-\infty}^{\infty}\mathrm{d}S_{XX}^{D}(f)$ 的归一化，以得到式(8-100)中的上极限 1。虽然两个密度函数

$f_X(x)$和$S_{XX}(f)$是非负实数，它们有一个重要区别，功率谱密度必须是偶函数。这意味着功率谱分布函数具有另外的性质：

- 中心归一化。$S_{XX}^D(0)/R_{XX}(0)=1/2$

如果$S_{XX}^D(f)$连续，那么$dS_{XX}^D(f)=S_{XX}(f)df$。如果$S_{XX}^D(f)$有不连续点，δ函数可以用于表示$S_{XX}(f)$(类似于表示离散和混合随机变量的$f_X(x)$)。

例 8-5 从表 8-2 发现，$R_{XX}(\tau)=\beta\exp(-\alpha|\tau|)$的$S_{XX}(f)=2\beta\alpha/(4\pi^2f^2+\alpha^2)$。对应的功率谱分布函数是：

$$\begin{aligned}S_{XX}^D(f)&=2\beta\alpha\int_{-\infty}^{f}\frac{1}{4\pi^2f^2+\alpha^2}df\\&=(\beta/\pi)\arctan(2\pi f/\alpha)|_{-\infty}^{f}=(\beta/\pi)\arctan(2\pi f/\alpha)+\beta/2\end{aligned}\qquad(8\text{-}102)$$

观察发现$S_{XX}^D(\infty)=\beta$，而β刚好是$R_{XX}(0)=\beta$的随机过程的功率。分别令$\beta=1$，$\alpha=2$，得到$S_{XX}(f)$和$S_{XX}^D(f)$，如图 8-8 所示。

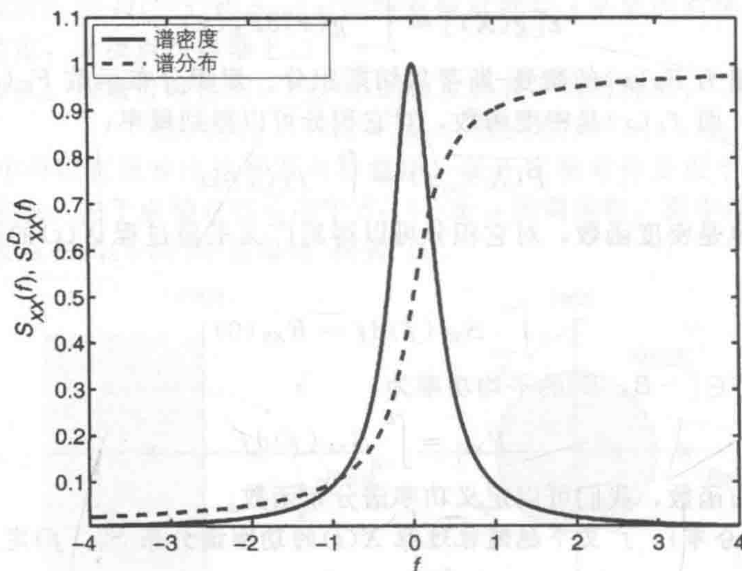


图 8-8 例 8-5 中的功率谱密度和功率谱分布，其中$\alpha=2$，$\beta=1$ ◀

例 8-6 假设谱分布分段线性：

$$S_{XX}^D(f)=\begin{cases}0, & f<-f_1\\ f/f_1+1, & |f|\leqslant f_1\\ 2, & f>f_1\end{cases}\qquad(8\text{-}103)$$

其中，$f_1>0$，相应的功率谱密度为：

$$S_{XX}(f)=\begin{cases}0, & |f|>f_1\\ 1/f_1, & |f|\leqslant f_1\end{cases}\qquad(8\text{-}104)$$

这是一个面积为 2 的矩形：$S_{XX}(f)=(1/f_1)\mathrm{rect}(f/2f_1)$，因此自相关函数为

$$R_{XX}(\tau)=2\mathrm{sinc}(2f_1\tau)\qquad(8\text{-}105)$$

总平均功率$R_{XX}(0)=2$，与$S_{XX}^D(\infty)$相等。 ◀

8.7 互功率谱密度

随机过程$X(t)$和$Y(t)$的互功率谱密度的计算与$S_{XX}(\omega)$的计算方法相同。

由帕斯瓦尔定理，功率可以表示为：

$$P_{XY}=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_{XY}(\omega)d\omega\qquad(8\text{-}106)$$

其中：

$$S_{XY}(\omega) \triangleq \lim_{T\to\infty}(1/2T)\varepsilon[X(\omega)Y^*(\omega)] \tag{8-107}$$

代入傅里叶变换得到：

$$S_{XY}(\omega) = \lim_{T\to\infty}\frac{1}{2T}\mathcal{E}\left[\int_{-T}^{T}X(t_1)\exp(-\mathrm{j}\omega t_1)\mathrm{d}t_1\int_{-T}^{T}Y(t_2)\exp(\mathrm{j}\omega t_2)\mathrm{d}t_2\right]$$

$$= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T}R_{XY}(t_1,t_2)\exp(-\mathrm{j}\omega(t_1-t_2))\mathrm{d}t_1\mathrm{d}t_2 \tag{8-108}$$

从 $S_{XX}(\omega)$推导过程可以看出，$S_{XY}(\omega)$的计算步骤大致相同，定义如下。

定义（互功率谱密度）　广义平稳过程 $X(t)$和 $Y(t)$，假设互相关函数 $R_{XY}(\tau)\triangleq\mathcal{E}[X(t)Y(t+\tau)]$，则互功率谱密度为：

$$S_{XY}(\omega) = \int_{-\infty}^{\infty}R_{XY}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau \tag{8-109}$$

其中 ω 是角频率。

（注意，一些书定义互相关函数为$\mathcal{E}[X(t+\tau)Y(t)]$。）自变量为 f 和 s 的其他形式的功率谱密度，以及那些自变量为 $\mathrm{j}\omega$ 和 z 的随机序列的功率谱密度，也可以用于定义互功率谱密度。

随机过程$\{X(t), Y(t)\}$的互功率谱密度具有下列性质。离散时间序列$\{X[k], Y[k]\}$的互功率谱密度 $S_{XY}(\mathrm{j}\omega)$具有类似的性质。

- 下标非对称性。$S_{YX}(\omega)=S_{XY}(-\omega)=S_{XY}^*(\omega)$。由互相关性质 $R_{YX}(\tau)=R_{XY}(-\tau)$：

$$S_{YX}(\omega) = \int_{-\infty}^{\infty}R_{YX}(\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau = \int_{-\infty}^{\infty}R_{XY}(-\tau)\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau \tag{8-110}$$

做变量替换 $\alpha=-\tau$，得到：

$$S_{YX}(\omega) = \int_{-\infty}^{\infty}R_{XY}(\alpha)\exp(-\mathrm{j}(-\omega)\alpha)\mathrm{d}\alpha = S_{XY}(-\omega) \tag{8-111}$$

由于$\{X(t), Y(t)\}$为实数，上式与 $S_{XY}^*(\omega)$相同。

- 偶部分。$\mathrm{Re}[S_{XY}(\omega)]$是偶函数。从 $S_{XY}(\omega)$和 $S_{YX}(\omega)$的定义：

$$\mathrm{Re}[S_{XY}(\omega)] = \int_{-\infty}^{\infty}R_{XY}(\tau)\cos(\omega\tau)\mathrm{d}\tau = \mathrm{Re}[S_{YX}(\omega)] \tag{8-112}$$

其中利用了 $\cos(-\omega\tau)=\cos(\omega\tau)$。再利用第一个性质，可以得到 $\mathrm{Re}[S_{XY}(\omega)]=\mathrm{Re}[S_{XY}(-\omega)]$，这是偶函数的定义。

- 奇部分。$\mathrm{Im}[S_{XY}(\omega)]$为奇函数。使用类似前面证明的方法：

$$\mathrm{Im}[S_{XY}(\omega)] = -\int_{-\infty}^{\infty}R_{XY}(\tau)\sin(\omega\tau)\mathrm{d}\tau = -\mathrm{Im}[S_{YX}(\omega)] \tag{8-113}$$

其中利用了 $\sin(-\omega\tau)=-\sin(\omega\tau)$。再利用第一个性质，可以得到 $\mathrm{Im}[S_{XY}(\omega)]=-\mathrm{Im}[S_{XY}(-\omega)]$，这是奇函数的定义。

- 正交性。若 $X(t)$和 $Y(t)$正交，则 $R_{XY}(\tau)=0$，$S_{XY}(\omega)=0$，由定义得到对于所有 τ，$R_{XY}(\tau)=0$，$S_{XY}(\omega)=0$。
- 不相关。若 $X(t)$和 $Y(t)$不相关，则 $R_{XY}(\tau)=\mathcal{E}[X(t)]\mathcal{E}[Y(t+\tau)]$，那么 $S_{XY}(\omega)=2\pi\mu_X\mu_Y\delta(\omega)$。因为过程不相关，$R_{XY}(\tau)=\mathcal{E}[X(t)Y(t+\tau)]=\mu_X\mu_Y$ 是一个常数。因此：

$$S_{XY}(\tau) = \mu_X\mu_Y\int_{-\infty}^{\infty}\exp(-\mathrm{j}\omega\tau)\mathrm{d}\tau = 2\pi\mu_X\mu_Y\delta(\omega) \tag{8-114}$$

这是由于常数的傅里叶变换是狄拉克 δ 函数。普通频率的相应结果是：

$$S_{XY}(f) = \mu_X\mu_Y\delta(f) \tag{8-115}$$

如果任一过程具有零均值，这往往是工程中的实际情况，互功率谱密度是零。

我们也可以定义互功率谱密度分布。

定义（互功率谱分布）　**互功率谱分布** $S_{XY}^D(f)$为

$$R_{XY}(\tau)=\int_{-\infty}^{\infty}\exp(\mathrm{j}2\pi f\tau)\mathrm{d}S_{XY}^{D}(f) \tag{8-116}$$

其中 $R_{XY}(\tau)$ 是随机过程 $\{X(t),\ Y(t)\}$ 的互相关函数。

互功率谱分布具有与 $S_{XX}^{D}(f)$ 相同的三个性质，并且如果它可导，则：

$$\mathrm{d}S_{XY}^{D}(f)=S_{XY}(f)\mathrm{d}f \tag{8-117}$$

例 8-7 在 8.8 节中，我们将考虑输入为随机信号的一些系统。这里，我们讨论滑动平均(Moving Average，MA)滤波器(本章后面也将介绍)的一个简单例子。假设自相关函数 $R_{XX}[m]=\delta[m]$ 的随机序列 $X[k]$ 通过脉冲响应 $h[k]=\delta[k]+\delta[k-1]$ 的滤波器，输出序列 $Y[k]=X[k]+X[k-1]$。互相关函数为：

$$\begin{aligned}R_{XY}[m]&=\mathcal{E}[X[k]X[k+m]]+\mathcal{E}[X[k]X[k+m-1]]\\&=\delta[m]+\delta[m-1]\end{aligned} \tag{8-118}$$

互功率谱密度为：

$$\begin{aligned}S_{XY}(\mathrm{j}\omega)&=\sum_{m=-\infty}^{\infty}(\delta[m]+\delta[m-1])\exp(-\mathrm{j}\omega m)\\&=1+\exp(-\mathrm{j}\omega)=2\exp(-\mathrm{j}\omega/2)\cos(\omega/2)\end{aligned} \tag{8-119}$$

观察发现

$$\mathrm{Re}[S_{XY}(\mathrm{j}\omega)]=1+\cos(\omega),\mathrm{Im}[S_{XY}(\mathrm{j}\omega)]=\sin(\omega) \tag{8-120}$$

分别是偶函数和奇函数。我们也可以写成

$$\begin{aligned}R_{YX}[m]&=\mathcal{E}[X[k]X[k+m]]+\mathcal{E}[X[k-1]X[k+m]]\\&=\delta[m]+\delta[m+1]\end{aligned} \tag{8-121}$$

因此：

$$S_{YX}(\mathrm{j}\omega)=1+\exp(\mathrm{j}\omega)=2\exp(\mathrm{j}\omega/2)\cos(\omega/2) \tag{8-122}$$

从中可以得出 $S_{YX}(\mathrm{j}\omega)=S_{XY}(-\mathrm{j}\omega)=S_{XY}^{*}(\omega)$。◀

在许多工科课程中，因为信号与系统假定为因果的，所以只研究了单边变换。然而，由于自相关函数是偶函数，功率谱密度的定义利用了双边变换。附录 C 的几个表中给出了傅里叶变换、离散时间傅里叶变换、拉普拉斯变换和 z 变换。在 8.8 节中，我们将探讨经过因果线性时不变系统滤波后，随机过程 $X(t)$ 的功率谱密度是如何改变的。因此，单边和双边变换的组合将用于研究滤波后的信号。当然，我们可以一直使用双边变换，在积分中包含单位阶跃函数以获得单边的结果。所以一定要谨慎利用表中的双边变换对(以及是否将时域函数乘以 $u(t)$)。

8.8　输入为随机信号的系统

虽然许多系统是多输入和多输出的，但本章侧重研究单输入和单输出(single-input single-output，SISO)系统，如图 8-9a 所示。假设输入是已知性能的随机过程 $X(t)$，例如它的一阶矩和二阶矩，或者它的概率密度函数以及是否相关。通常情况下，系统分析或设计的目的是改变信号 $X(t)$，以产生满足要求的输出 $Y(t)$。例如，$X(t)$ 可能是包含加性噪声的通信信号，系统的目的是对信号 $X(t)$ 进行滤波以减少噪声，并且不影响感兴趣的潜在信号。图 8-9b 给出了均衡器的结构，其目的是补偿发射序列 $X[k]$ 通过信道 $H(z)$ 后的失真。图 8-9c 给出了系统辨识的结构，其中 $G(z)$ 为 $H(z)$ 的一个模型，通过输入序列 $X[k]$ 和输出序列 $Y[k]$ 的测量结果得到。当然，在具体实现中，各种滤波器针对的是随机序列的现实。像前面章节那样，我们用概率表示来描述如何处理现实的合集。

一般地，系统可以分类为：(ⅰ)线性和非线性；(ⅱ)不随时间变化和随时间变化。线性系统满足叠加性，因此如果确定性信号 $x_1(t)$ 和 $x_2(t)$ 对应的输出分别为 $y_1(t)$ 和 $y_2(t)$ 的，那么输入 $c_1x_1(t)+c_2x_2(t)$ 对应的输出为 $c_1y_1(t)+c_2y_2(t)$，其中 $\{c_n\}$ 是常数。虽然大多数物理系统非线性且难以分析和控制，但很多可以用线性模型合理地近似，尽管可能只在有限的动态范围和有限的持续时间内。在本章中，我们只研究线性时不变系统，因为它

们可以由冲激响应函数描述，且输出由卷积获得(第1章回顾了这些连续和离散系统)。

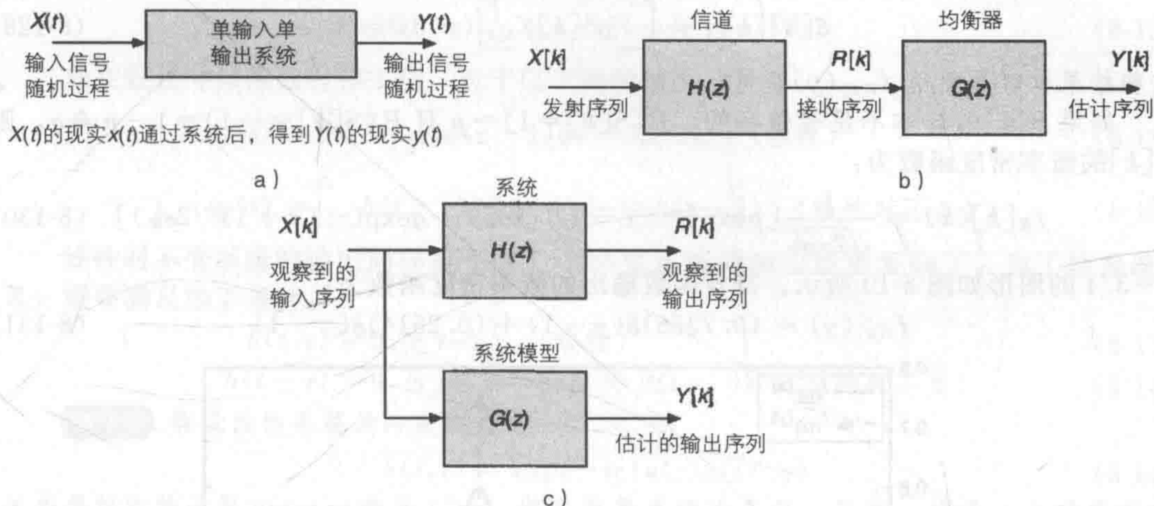


图 8-9 线性时不变系统。a) 单输入单输出系统；b) 信道均衡；c) 系统辨识

但是，我们首先简要地考虑无记忆非线性系统，在任何时刻的输出仅取决于当前时刻的输入。

8.8.1 非线性系统

无记忆非线性系统可以看作任意时刻输入的随机信号的变换或映射，因此

$$Y(t) = g(X(t)) \tag{8-123}$$

其中 $g(\cdot)$是一个确定性函数。工程中常见的例子如 $X^2(t)$和符号函数 $\mathrm{sgn}(X(t))$。由于系统无记忆，从而不存储过去时刻的输入值，可以应用第4章随机变量变换的方法。如果函数随时间变化，用符号表示就是 $g(X(t),t)$，只需要考虑当前变换以确定输出的特性。当然，如果变换变化，即使输入平稳，输出也是非平稳过程。

例 8-8 考虑随机序列 $X[k]$的非线性滤波器：

$$Y[k] = \mathrm{sgn}(X[k]) \tag{8-124}$$

并且假设：

$$X[k] = S[k] + W[k] \tag{8-125}$$

其中 $S[k]$是样本为±1的独立同分布的对称伯努利序列，$W[k]$是方差为 σ_W^2 的加性高斯白噪声(Additive White Gaussian Noise，AWGN)序列。显然，$Y[k]$也是样本为±1的伯努利序列。假设 $S[k]$和 $W[k]$独立，从第4章知道 $X[k]$的概率密度函数为：

$$\begin{aligned} f_{X[k]}(x) &= f_{S[k]}(s) * f_{W[k]}(w) \\ &= \frac{1}{2\sqrt{2\pi\sigma_W^2}}\int_{-\infty}^{\infty}[\delta(x-1-v)+\delta(x+1-v)]\exp(-v^2/2\sigma_W^2)\mathrm{d}v \\ &= \frac{1}{2\sqrt{2\pi\sigma_W^2}}[\exp(-(x-1)^2/2\sigma_W^2)+\exp(-(x+1)^2/2\sigma_W^2)] \end{aligned} \tag{8-126}$$

图4-33给出了这种双峰概率密度函数的例子。滤波器输出序列的概率密度函数为

$$f_{Y[k]}(y) = (1/2)[\delta(y-1)+\delta(y+1)] \tag{8-127}$$

这种类型的非线性滤波器是通信检测的一个例子：传输信号1或−1受高斯噪声影响，对应的概率密度函数如图4-33所示。当±1等概率时，$\mathrm{sgn}(X[k])$是从 $X[k]$中估计 $S[k]$的最优决策器。(第10章将进一步讨论信号检测)。不同时刻 $Y[k]$的矩可以直接利用 $X[k]$的概率密度函数计算：

$$\mathcal{E}[Y[k]] = \int_{-\infty}^{\infty}\mathrm{sgn}(x)f_{X[k]}(x)\mathrm{d}x = 0 \tag{8-128}$$

这是因为 sgn(·)是奇函数，$X[k]$的概率密度函数为偶函数。又因为 $Y^2[k]=1$，

$$\mathcal{E}[Y^2[k]] = \int_{-\infty}^{\infty} y^2[k] f_{X[k]}(x)\mathrm{d}x = 1 \tag{8-129}$$

这些结果也容易使用 $f_{Y[k]}(y)$获得。◀

如果 $S[k]$的样本不是等概率的，$P(S[k]=1)=p$ 且 $P(S[k]=-1)=1-p\triangleq q$，则 $X[k]$的概率密度函数为：

$$f_X[k](x) = \frac{1}{\sqrt{2\pi\sigma_W^2}}[p\exp(-(x-1)^2/2\sigma_W^2) + q\exp(-(x+1)^2/2\sigma_W^2)] \tag{8-130}$$

$p=3/4$ 的图形如图 8-10 所示。符号函数输出的概率密度函数为：

$$f_{Y[k]}(y) = (0.7386)\delta(y-1) + (0.2614)\delta(y+1) \tag{8-131}$$

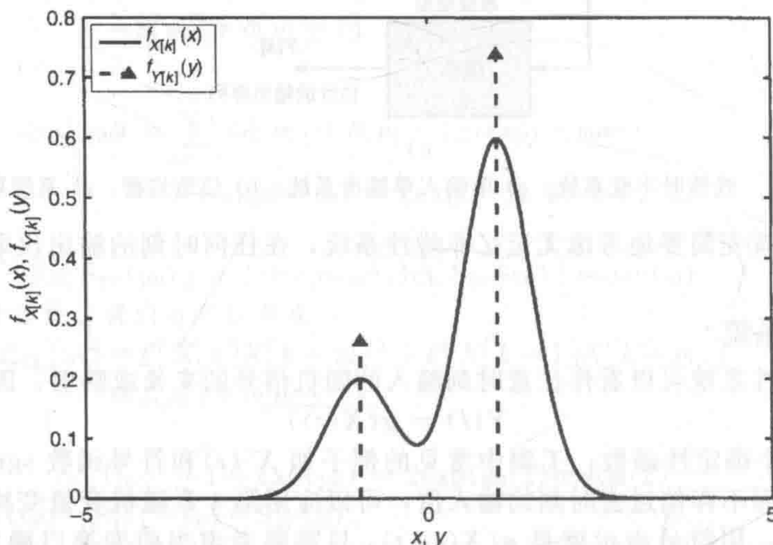


图 8-10 例 8-8 中对称伯努利序列的概率密度函数，其中样本为±1，加性高斯白噪声的方差为 1，$p=3/4$

其中由于 $p\neq q$：

$$\int_0^{\infty} f_{X[k]}(x)\mathrm{d}x = 0.7386,\quad \int_{-\infty}^{0} f_{X[k]}(x)\mathrm{d}x = 0.2614 \tag{8-132}$$

输出具有以下矩：

$$\mathcal{E}[Y[k]] = 0.4772,\quad \mathcal{E}[Y^2[k]] = 1 \tag{8-133}$$

对于这种情况，由于 $X[k]$的概率密度函数不再是偶函数，符号函数不是最佳检测器。如果 $p>1/2$，$S[k]=1$ 更可能是发送信号，接收机需要考虑这种情况。这意味着，符号函数检测器与 0 的交点应向原点左边移动，平移量取决于 p 的值。

第 12 章讨论了其他非线性系统的例子。

8.8.2 线性系统

对于不一定时不变的线性系统，滤波器输出的随机过程 $Y(t)$通过下列积分得到：

$$Y(t) = \int_{-\infty}^{t} h(t,\tau)X(\tau)\mathrm{d}\tau \tag{8-134}$$

其中 $X(t)$是输入的随机过程，$h(t,\tau)$ 是该系统的确定性冲激响应函数。由线性系统的叠加性得到时变系统的这个积分：$h(t,\tau)$ 分别取决于 t 和 τ。对于线性时不变系统，$h(t,\tau)=h(t-\tau)$ 仅取决于时间差，而不是分别与 t 和 τ 有关。冲激响应函数通过令输入为狄拉克 δ 函数，再利用取样特性得到：

$$y(t) = \int_{-\infty}^{t} h(t-\tau)\delta(\tau)\mathrm{d}\tau = h(t) \tag{8-135}$$

由冲激函数 $\delta(\tau)$可以得出冲激出现的时刻为 $\tau=0$，进而得到 $h(t-\tau)$在该时刻的取值。

对于时变系统，冲激响应是

$$y(t)=\int_{-\infty}^{t}h(t,\tau)\delta(\tau)\mathrm{d}\tau=h(t,0) \tag{8-136}$$

现在假设冲激函数右移 1 位，对于以下两种情况：

$$y(t)=\int_{-\infty}^{t}h(t,\tau)\delta(\tau-1)\mathrm{d}\tau=h(t,1)\quad（线性） \tag{8-137}$$

$$y(t)=\int_{-\infty}^{t}h(t-\tau)\delta(\tau-1)\mathrm{d}\tau=h(t-1)\quad（线性时不变） \tag{8-138}$$

线性时不变系统的输出同样有平移，而时变系统的输出就更复杂了。为了使系统因果，需要满足如下条件：

$$h(t,\tau)=0,当\ t<\tau\quad（线性） \tag{8-139}$$

$$h(t-\tau)=0,当\ t<\tau\Rightarrow h(t)=0(t<0)\quad（线性时不变） \tag{8-140}$$

例 8-9 假设线性系统的冲激响应为

$$h(t,\tau)=\exp(-t\tau)u(\tau)u(t-\tau) \tag{8-141}$$

其中单位阶跃函数 $u(t-\tau)$ 确保 $t>\tau$，满足因果系统的条件。显然，这是一个时变系统，因为随着 τ 的改变系统响应发生变化。对于线性时不变系统，有：

$$h(t-\tau)=\exp(-(t-\tau))u(\tau)u(t-\tau) \tag{8-142}$$

其形状不随 τ 的变化而变化。假设系统的输入为 $x(t)=\exp(-t)u(t)$ 对于时变系统：

$$\begin{aligned}y(t)&=\int_{0}^{t}\exp(-t\tau)\exp(-\tau)\mathrm{d}\tau\\&=\int_{0}^{t}\exp(-\tau(t+1))\mathrm{d}\tau=\frac{1}{t+1}[1-\exp(-t^2-t)]u(t)\end{aligned} \tag{8-143}$$

线性时不变系统相应的结果是：

$$\begin{aligned}y(t)&=\int_{0}^{t}\exp(-(t-\tau))\exp(-\tau)\mathrm{d}\tau\\&=\int_{0}^{t}\exp(-t)\mathrm{d}\tau=t\exp(-t)u(t)\end{aligned} \tag{8-144}$$

如果输入有时移 $x(t)=\exp(-(t-1))u(t-1)$，时变系统的输出是：

$$\begin{aligned}y(t)&=\int_{1}^{t}\exp(-t\tau)\exp(-(\tau-1))\mathrm{d}\tau\\&=\frac{\exp(1)}{t+1}[\exp(-t-1)-\exp(-t^2-t)]u(t-1)\\&=\frac{1}{t+1}[\exp(-t)-\exp(-t^2-t+1)]u(t-1)\end{aligned} \tag{8-145}$$

而对于线性时不变系统：

$$\begin{aligned}y(t)&=\int_{1}^{t}\exp(-(t-\tau))\exp(-(\tau-1))\mathrm{d}\tau\\&=\int_{1}^{t}\exp(-(t-1))\mathrm{d}\tau=(t-1)\exp(-(t-1))u(t-1)\end{aligned} \tag{8-146}$$

图 8-11 给出两个系统的结果，可以看出，时移后的输入信号经过时变系统后，产生了更为复杂的输出信号，而不像线性时不变系统那样只进行了简单的时移。◀

接下来，我们讨论输入为随机信号时，线性时不变系统的信号表示。系统输入、输出的互相关函数是：

$$\begin{aligned}R_{XY}(\tau)&=\mathcal{E}[X(t)Y(t+\tau)]\\&=\mathcal{E}\left[X(t)\int_{-\infty}^{\infty}h(t+\tau-\alpha)X(\alpha)\mathrm{d}\alpha\right]\end{aligned} \tag{8-147}$$

是均方意义下的积分。交换积分和期望顺序：

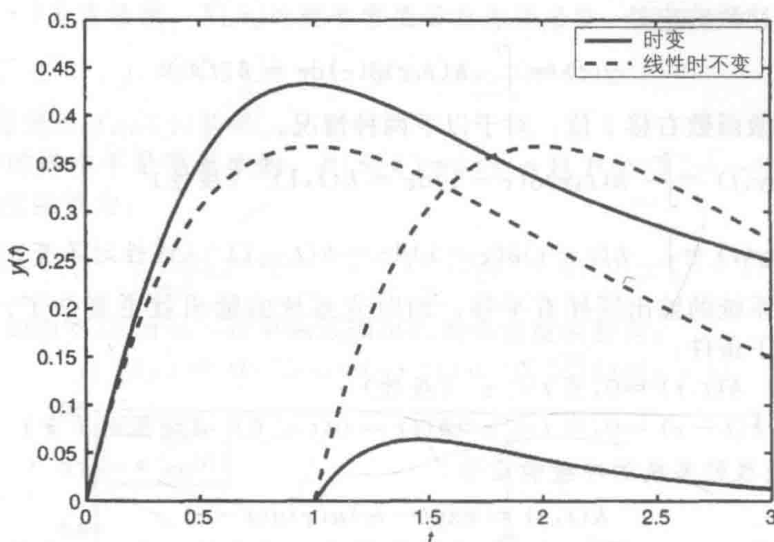


图 8-11 例 8-9 中时变和线性时不变系统的输出波形。对于时变系统，输入信号的时移不能产生输出信号的简单时移

$$R_{XY}(\tau)=\int_{-\infty}^{\infty}h(t+\tau-\alpha)\,\mathcal{E}[X(t)X(\alpha)]\mathrm{d}\alpha=\int_{-\infty}^{\infty}h(t+\tau-\alpha)R_{XX}(\alpha-t)\mathrm{d}\alpha \tag{8-148}$$

这里假设 $X(t)$广义平稳。定义 $v\triangleq\alpha-t$，整理上式得到：

$$R_{XY}(\tau)=\int_{-\infty}^{\infty}h(\tau-v)R_{XX}(v)\mathrm{d}v \tag{8-149}$$

它是 $R_{XX}(\tau)$和 $h(\tau)$的卷积：

$$R_{XY}(\tau)=R_{XX}(\tau)*h(\tau) \tag{8-150}$$

同样地，我们可以证明

$$R_{YX}(\tau)=R_{XX}(\tau)*h(-\tau) \tag{8-151}$$

及(见习题 8-17)

$$R_{YY}(\tau)=R_{XX}(\tau)*h(\tau)*h(-\tau) \tag{8-152}$$

$R_{XY}(\tau)$的功率谱密度为：

$$\begin{aligned}S_{XY}(f)&=\int_{-\infty}^{\infty}R_{XY}(\tau)\exp(-\mathrm{j}2\pi f\tau)\mathrm{d}\tau\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\tau-\alpha)R_{XX}(\alpha)\exp(-\mathrm{j}2\pi f\tau)\mathrm{d}\alpha\mathrm{d}\tau\end{aligned} \tag{8-153}$$

代入 $t=\tau-\alpha$，得到：

$$\begin{aligned}S_{XY}(f)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(t)R_{XX}(\alpha)\exp(-\mathrm{j}2\pi f(t+\alpha))\mathrm{d}\alpha\mathrm{d}t\\&=\int_{-\infty}^{\infty}h(t)\exp(-\mathrm{j}2\pi ft)\mathrm{d}t\int_{-\infty}^{\infty}R_{XX}(\alpha)\exp(-\mathrm{j}2\pi f\alpha)\mathrm{d}\alpha\\&=H(f)S_{XX}(f)\end{aligned} \tag{8-154}$$

这里利用了 $S_{XX}(f)$为偶函数。同样，

$$S_{YX}(f)=H(-f)S_{XX}(f) \tag{8-155}$$

及(见习题 8-18)

$$S_{YY}(f)=H(f)H(-f)S_{XX}(f) \tag{8-156}$$

对于实冲激响应函数，$H(-f)=H^*(f)$，从而

$$S_{YY}(f)=|H(f)|^2S_{XX}(f) \tag{8-157}$$

表 8-3 总结了角频率 ω 以及使用 $H(s)$的相应结果。同样包括了使用 $H(\mathrm{j}\omega)$和 $H(z)$的随机序列的相应结果。

表 8-3 线性时不变系统和概率密度函数

连续时间系统	离散时间系统
$S_{YY}(\omega)=\|H(\omega)\|^2 S_{XX}(\omega)$	$S_{YY}(\mathrm{j}\omega)=\|H(\mathrm{j}\omega)\|^2 S_{XX}(\mathrm{j}\omega)$
$S_{YY}(s)=H(s)H(-s)S_{XX}(s)$	$S_{YY}(z)=H(z)H(z^{-1})S_{XX}(z)$
$S_{XY}(\omega)=H(\omega)S_{XX}(\omega)$	$S_{XY}(\mathrm{j}\omega)=H(\mathrm{j}\omega)S_{XX}(\mathrm{j}\omega)$
$S_{XY}(s)=H(s)S_{XX}(s)$	$S_{XY}(z)=H(z)S_{XX}(z)$
$S_{YX}(\omega)=H(-\omega)S_{XX}(\omega)$	$S_{YX}(\mathrm{j}\omega)=H(-\mathrm{j}\omega)S_{XX}(\mathrm{j}\omega)$
$S_{YX}(s)=H(-s)S_{XX}(s)$	$S_{YX}(z)=H(z^{-1})S_{XX}(z)$

例 8-10 在这个例子中，我们演示了均值为 0，方差为 1 的不相关高斯序列的滤波效果，如图 8-12a 所示。滤波器的冲激响应为：

$$h[k]=\frac{1}{N}\sum_{m=0}^{N-1}\delta[k-m] \tag{8-158}$$

其中 $N=3$。滑动平均滤波器具有有限脉冲响应(FIR)(见第 1 章)，传递函数为：

$$H(\mathrm{j}\omega)=(1/3)\sum_{k=0}^{2}\exp(-\mathrm{j}\omega k)=(1/3)\exp(-\mathrm{j}\omega)(\exp(\mathrm{j}\omega)+1+\exp(-\mathrm{j}\omega))$$

$$=(1/3)\exp(-\mathrm{j}\omega)[1+2\cos(\omega)] \tag{8-159}$$

由于 $S_{XX}(\mathrm{j}\omega)=1$，输出的功率谱密度为：

$$S_{YY}(\mathrm{j}\omega)=|H(\mathrm{j}\omega)|^2=(1/9)[1+2\cos(\omega)]^2 \tag{8-160}$$

自相关函数等于：

$$R_{YY}[m]=(1/9)(3\delta[m]+2\delta[m-1]+2\delta[m+1]+\delta[m-2]+\delta[m+2]) \tag{8-161}$$

它也很容易通过 $h[k]$ 和 $h[-k]$ 的卷积获得。图 8-13 画出了 $S_{YY}(\mathrm{j}\omega)$ 和 $R_{YY}[m]$。图 8-12b 给出了一个滤波器的输出现实，显而易见，该输出是相关的。

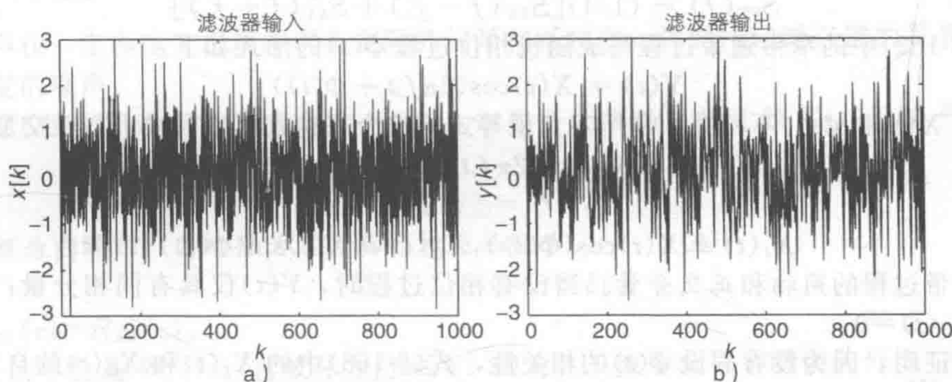


图 8-12 例 8-10 中线性时不变系统的现实。a) 不相关的高斯输入 $X[k]$；b) 滤波器的相关输出 $Y[k]$

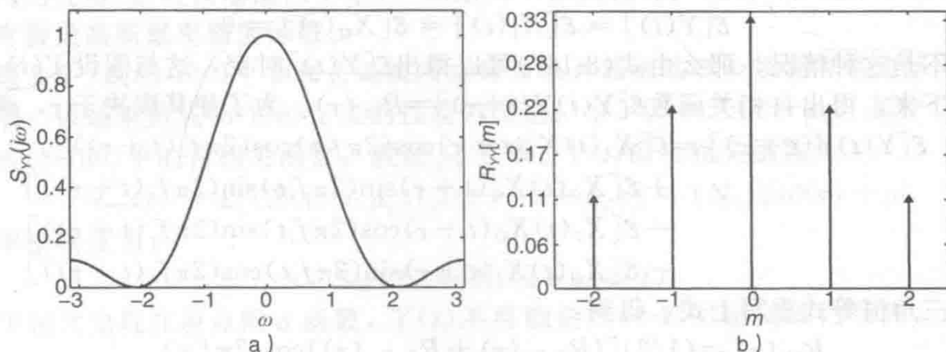


图 8-13 例 8-10 中不相关高斯序列的滤波结果。a) 滤波器输出的功率谱密度；b) 滤波器输出的自相关函数

8.9 通带信号

到目前为止，我们主要研究了功率谱密度在原点附近分布的随机信号，这些信号也叫做基带信号。在大多数通信系统中，感兴趣信号的功率谱密度在中心频率 f_c 附近分布，其中 $f_c \gg 0$ 这样的信号称为通带信号。图 8-14画出了基带和通带信号的 $S_{XX}(f)$。观察发现基带信号的单边带宽为 $B(f \geqslant 0)$，而通带信号的单边带宽为 $2B$。这是因为通带信号通常是基带信号经过载频为 f_c 的正弦信号调制而成。当通带信号的带宽 $B \ll f_c$时，称该信号为窄带通带信号。

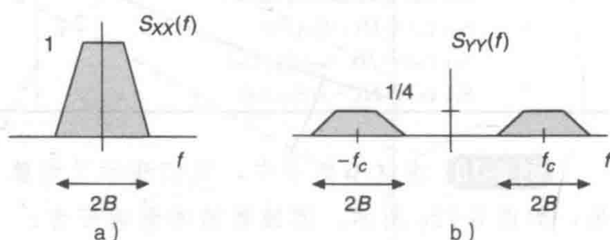


图 8-14　功率谱密度。a）基带随机过程 $X(t)$；b）通过对 $X(t)$进行调制，得到通带随机过程 $Y(t)$

设 $X(t)$为功率谱密度为 $S_{XX}(f)$的基带随机过程，其单边带宽为 B。假设 $X(t)$乘以载波频率为 f_c 的正弦波：

$$Y(t) = X(t)\cos(2\pi f_c t) \tag{8-162}$$

为确定 $S_{YY}(f)$，得出自相关函数：

$$R_{YY}(\tau) = \mathcal{E}[X(t)\cos(2\pi f_c t)X(t+\tau)\cos(2\pi f_c(t+\tau))] \tag{8-163}$$

在$[-\pi, \pi]$上均匀分布的随机相位 Φ 通常包括在余弦的自变量中。工程中采用这种方法建模时，余弦不一定对齐随机过程的起始时间。假定随机相位和随机过程不相关，得到：

$$\begin{aligned}R_{YY}(\tau) &= \mathcal{E}[X(t)X(t+\tau)]\,\mathcal{E}[\cos(2\pi f_c t+\Phi)\cos(2\pi f_c(t+\tau)+\Phi)]\\ &= (1/2)R_{XX}(\tau)\,\mathcal{E}[\cos(2\pi f_c\tau)+\cos(2\pi f_c(2t+\tau)+2\Phi)]\\ &= (1/2)R_{XX}(\tau)\cos(2\pi f_c\tau)\end{aligned} \tag{8-164}$$

其中利用了余弦在随机相位上的平均值是零的条件。利用傅里叶变换的调制性质(见附录 C)：

$$S_{YY}(f) = (1/4)[S_{XX}(f-f_c)+S_{XX}(f+f_c)] \tag{8-165}$$

式(8-162)中的窄带通带过程写成随机相位过程 $\Phi(t)$的形式如下：

$$Y(t) = X(t)\cos(2\pi f_c t+\Phi(t)) \tag{8-166}$$

这里假设 $X(t)$和 $\Phi(t)$不相关。利用三角恒等式，这个表达式可以写成下列正交形式：

$$Y(t) = X_I(t)\cos(2\pi f_c t) - X_Q(t)\sin(2\pi f_c t) \tag{8-167}$$

其中

$$X_I(t) \triangleq X(t)\cos(\Phi(t)),\ X_Q(t) \triangleq X(t)\sin(\Phi(t)) \tag{8-168}$$

分别是通带过程的同相和正交分量。当没有相位过程时，$Y(t)$仅具有同相分量：$X_I(t)=X(t)$，$X_Q(t)=0$。

事实证明，因为没有假设 $\Phi(t)$的相关性，式(8-168)中的 $X_I(t)$和 $X_Q(t)$的自相关函数(与功率谱密度)不容易得到。然而，假定 $Y(t)$广义平稳，那么可以得出 $X_I(t)$和 $X_Q(t)$的一些性质。首先，显然

$$\mathcal{E}[Y(t)] = \mathcal{E}[X_I(t)] = \mathcal{E}[X_Q(t)] = 0 \tag{8-169}$$

如果不是这种情况，那么由式(8-167)可以得出$\mathcal{E}[Y(t)]$时变，这与假设 $Y(t)$广义平稳相悖。接下来，得出自相关函数$\mathcal{E}[Y(t)Y(t+\tau)]=R_{YY}(\tau)$。为了使其取决于 τ，观察

$$\begin{aligned}\mathcal{E}[Y(t)Y(t+\tau)] &= \mathcal{E}[X_I(t)X_I(t+\tau)\cos(2\pi f_c t)\cos(2\pi f_c(t+\tau))]\\ &\quad+\mathcal{E}[X_Q(t)X_Q(t+\tau)\sin(2\pi f_c t)\sin(2\pi f_c(t+\tau))]\\ &\quad-\mathcal{E}[X_I(t)X_Q(t+\tau)\cos(2\pi f_c t)\sin(2\pi f_c(t+\tau))]\\ &\quad-\mathcal{E}[X_Q(t)X_I(t+\tau)\sin(2\pi f_c t)\cos(2\pi f_c(t+\tau))]\end{aligned} \tag{8-170}$$

利用三角恒等式重写上式，得到：

$$\begin{aligned}R_{YY}(\tau) &= (1/2)[(R_{X_IX_I}(\tau)+R_{X_QX_Q}(\tau))\cos(2\pi f_c\tau)\\ &\quad+(R_{X_IX_I}(\tau)-R_{X_QX_Q}(\tau))\cos(2\pi f_c(2t+\tau))\end{aligned}$$

$$-(R_{X_QX_I}(\tau)-R_{X_IX_Q}(\tau))\sin(2\pi f_c\tau)$$
$$-(R_{X_QX_I}(\tau)+R_{X_IX_Q}(\tau))\sin(2\pi f_c(2t+\tau))] \tag{8-171}$$

因为$Y(t)$广义平稳，包含t的各项必须为0，得到相关函数的下列条件：

$$R_{X_IX_I}(\tau)=R_{X_QX_Q}(\tau) \tag{8-172}$$
$$R_{X_IX_Q}(\tau)=-R_{X_QX_I}(\tau)=-R_{X_IX_Q}(-\tau) \tag{8-173}$$

其中最后一个等号对于任何互相关函数都成立。对于窄带通带过程，我们看到同相和正交分量的互相关必须是奇函数(一般不成立)。从而

$$R_{X_IX_Q}(0)=0 \tag{8-174}$$

有了这些条件，$R_{YY}(\tau)$也有一个正交形式：

$$R_{YY}(\tau)=R_{X_IX_I}(\tau)\cos(2\pi f_c\tau)-R_{X_QX_I}(\tau)\sin(2\pi f_c\tau) \tag{8-175}$$

观察发现，只有当$\tau=0$时，$R_{X_IX_I}(\tau)$、$R_{X_QX_Q}(\tau)$和$R_{YY}(\tau)$相等：

$$R_{X_IX_I}(0)=R_{X_QX_Q}(0)=R_{YY}(0) \tag{8-176}$$

因此，所有三个过程具有相同的方差(因为都是零均值)。我们也可以写出相应功率谱密度的表达式。利用傅里叶变换的调制性质：

$$S_{YY}(f)=(1/2)[S_{X_IX_I}(f-f_c)+S_{X_IX_I}(f+f_c)]$$
$$-(1/2\mathrm{j})[S_{X_QX_I}(f-f_c)-S_{X_QX_I}[f+f_c)] \tag{8-177}$$

因为自相关函数是一个正交信号，功率谱密度需要包含虚数项。从式(8-176)，三个功率谱密度的平均功率相等：

$$\int_{-\infty}^{\infty}S_{X_IX_I}(f)\mathrm{d}f=\int_{-\infty}^{\infty}S_{X_QX_Q}(f)\mathrm{d}f=\int_{-\infty}^{\infty}S_{YY}(f)\mathrm{d}f \tag{8-178}$$

第10章介绍通信系统时，将再次考虑通带信号。

8.10 白噪声

噪声在一定程度上出现在所有实际的信号与系统中。我们比较感兴趣的是具有平坦功率谱密度的噪声。

定义(白噪声过程)　对于任意$f\in\mathcal{R}$，如果广义平稳随机过程$X(t)$具有平坦功率谱密度

$$S_{XX}(f)=N_o/2 \tag{8-179}$$

称该过程为**白噪声**。其均值为零，其自相关函数为

$$R_{XX}(\tau)=(N_o/2)\delta(\tau) \tag{8-180}$$

因此$C_{XX}(\tau)=R_{XX}(\tau)$。

N_o的单位是W/Hz。它被称为白噪声，因为(ⅰ)波形有一个嘈杂的外观(在可听频率范围内听起来像噪声)；(ⅱ)所有频率相等(频谱平坦)，这类似于白色光在电磁频谱的可见范围内的性质(在波长范围内平坦)。此定义不涉及白噪声的分布，它可以是任何类型，尽管通常假设高斯概率密度函数。

定理 8-5　式(8-180)中自相关函数对应的白噪声随机过程必须具有零均值。

证明：该结果是式(8-179)定义的性质的结论。令$Y(t)=X(t)+\mu_Y$，其中$X(t)$具有零均值和式(8-180)中的自相关函数。假设μ_Y非零。$Y(t)$的自相关函数是

$$R_{YY}(\tau)=\mathcal{E}[(X(t)+\mu_Y)(X(t+\tau)+\mu_Y)]=(N_o/2)\delta(\tau)+\mu_Y^2 \tag{8-181}$$

其中功率谱密度为：

$$S_{YY}(f)=N_o/2+\mu_Y^2\delta(f) \tag{8-182}$$

由于包含出现在原点的δ函数，$Y(t)$不可能是式(8-179)定义的白噪声，除非它的均值是零。

第6章讨论过$X(t)$满足下列条件时，它是不相关的随机过程

$$R_{XX}(\tau)=\begin{cases}\mathcal{E}[X^2(t)], & \tau=0\\ \mathcal{E}[X(t)]\mathcal{E}[X(t+\tau)], & \tau\neq 0\end{cases} \tag{8-183}$$

与式(8-181)类似，它可以写成 $R_{XX}(\tau)=\alpha\delta(\tau)+\mu_X^2$。因此，具有平坦功率谱密度的白噪声意味着不相关的过程，但是反过来只有当均值为零时才成立。

高斯分布的白噪声过程可以通过对维纳过程求导得到：$X(t)=Y'(t)$，其中 $Y(t)$ 的自相关函数为：

$$R_{YY}(t_1,t_2)=c\min(t_1,t_2) \tag{8-184}$$

其中 $c=\alpha\varepsilon^2$(见第 6 章)。类似式(7-119)中对泊松过程的微分，我们可以求导得到维纳过程自相关函数：

$$R_{XX}(t_1,t_2)=\frac{\partial^2}{\partial t_1\partial t_2}R_{YY}(t_1,t_2)=c\delta(t_2-t_1) \tag{8-185}$$

这与式(8-180)相同。因此，尽管由于 $t_1=t_2$ 时 δ 函数的存在，均方微分技术上不存在，但是可以看出，维纳过程的微分是白噪声。

白噪声过程在实际中不能实现，因为它的平均功率为无穷大(等于方差)：

$$\sigma_X^2=C_{XX}(0)=R_{XX}(0)=\int_{-\infty}^{\infty}S_{XX}(f)\mathrm{d}f=\infty \tag{8-186}$$

然而，由于实际系统中有各种滤波器，$X(t)$ 通常是带限的，功率谱密度只在某些感兴趣频率范围内平坦，其余频率为 0：

$$S_{XX}(f)=\begin{cases}N_o/2, & |f|\leqslant B\\ 0, & \text{其他}\end{cases} \tag{8-187}$$

其中 B 是单边带宽。这种类型的功率谱密度对应的过程技术上不是白噪声，尽管当 B 远远大于感兴趣信号带宽时，它称为*带限白噪声*。由于 $f=0$ 时，$S_{XX}(f)$ 不等于 0，所以式(8-187)中的功率谱密度对应为基带信号。通带信号的功率谱密度的形式为：

$$S_{XX}(f)=\begin{cases}N_o/2, & |f|\in[f_c-B/2,f_c+B/2]\\ 0, & \text{其他}\end{cases} \tag{8-188}$$

其中 f_c 表示中心频率。图 8-15 给出了通带信号之一的调幅(Amplitude Modulated，AM)信号的频谱。

$$X(t)=M(t)\cos(2\pi f_c t)+N(t) \tag{8-189}$$

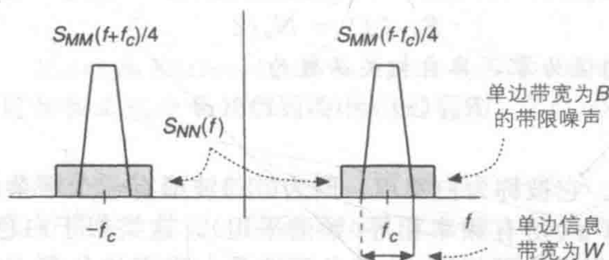


图 8-15 在带限白噪声中的 AM 信号的功率谱分量。$N(t)$ 的外观为白噪声，因为它的带宽 B 大于信息带宽 W。(注意 $S_{NN}(f)$ 与信号频谱 $[S_{MM}(f+f_c)+S_{MM}(f-f_c)]/4$ 分开显示。两个频谱相加得到总 $S_{XX}(f)$)

其中，$M(t)$ 是随机信息信号，f_c 为通带中心频率，$N(t)$ 是带限白噪声。利用前面通带信号的结论，可以直接得出：

$$S_{XX}(f)=(1/4)[S_{MM}(f-f_c)+S_{MM}(f+f_c)]+S_{NN}(f) \tag{8-190}$$

这里假设 $M(t)$ 和 $N(t)$ 不相关。

对于带限白噪声，方差为有限值：

$$\sigma_X^2=\int_{-B}^{B}S_{XX}(f)\mathrm{d}f=N_oB \tag{8-191}$$

前面定义了单边带宽 B，式(8-187)中功率谱密度 N_o 前面有因子 1/2，这是因为 σ_X^2 要

除以整个(双边)带宽 $2B$(用以简化表达式)。对于调幅信号，假定 $M(t)$具有零均值，如果随机信息的功率

$$P_M = R_{MM}(0) = \int_{-W}^{W} S_{MM}(f)\mathrm{d}f \tag{8-192}$$

那么接收信号的信噪比(Signal-to-Noise Ratio，SNR)是

$$\mathrm{SNR} = \frac{P_M}{N_o B} \tag{8-193}$$

信噪比是接收信号质量的一个测度：它决定了信息 $M(t)$能够可靠恢复的程度。观察发现随着 B 的增加信噪比降低，这是因为出现了更多的噪声。接收机前端滤波器的目标是在不损失信息的前提下尽可能多的去除噪声。这些滤波器的带宽 B 要尽可能接近信息带宽 W，而且在信息频率范围内的频率响应要保持相对平坦。

带限白噪声的自相关函数为：

$$\begin{aligned} R_{XX}(\tau) &= \int_{-B}^{B}(N_o/2)\exp(\mathrm{j}2\pi f\tau)\mathrm{d}f \\ &= (N_o/2)[\exp(\mathrm{j}2\pi B\tau) - \exp(-\mathrm{j}2\pi B\tau)]/(\mathrm{j}2\pi\tau) = N_o B\mathrm{sinc}(2B\tau) \end{aligned} \tag{8-194}$$

其中 $\mathrm{sinc}(\tau) \triangleq \sin(\pi\tau)/\pi\tau$ 是 sinc 函数。图 8-16 画出了不同噪声带宽对应的自相关函数，从中可以看出，主瓣宽度与 B 成反比。附录 B 中说明了

$$\lim_{B\to\infty} B\mathrm{sinc}(2\tau) = \delta(\tau) \tag{8-195}$$

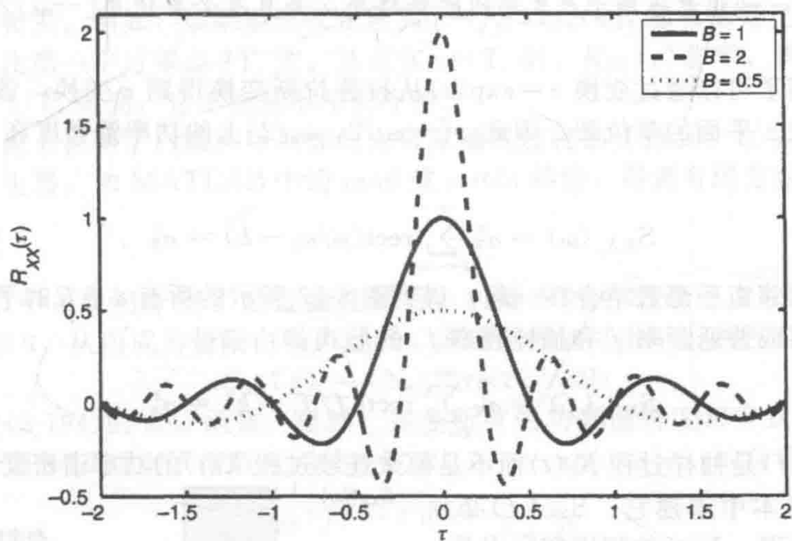


图 8-16　不同单边带宽 B 对应的带限白噪声的自相关函数

因此

$$\lim_{B\to\infty} N_o B\mathrm{sinc}(2B\tau) = (N_o/2)\delta(\tau) \tag{8-196}$$

这是无限带宽的功率谱密度的自相关函数。

接下来，我们讨论白噪声过程和白噪声序列的联系和区别。第 6 章中说明了随机序列 $X[k]$不相关的充分条件是

$$\begin{aligned} R_{XX}[m] &= \begin{cases} \mathcal{E}[X^2[k]], & m = 0 \\ \mathcal{E}[X[m]]\,\mathcal{E}[X[k+m]], & m \neq 0 \end{cases} \\ &= (\sigma_X^2 + \mu_X^2)\delta[m] + \mu_X^2 \end{aligned} \tag{8-197}$$

其中 σ_X^2 是有限方差。习题 8-13 证明了如果 $X[k]$是一个白噪声序列(功率谱密度平坦)，它的均值必须为 0，这与先前讨论的连续时间随机过程相同。假设均值为 0，那么 $R_{XX}[m] = \sigma_X^2\delta[m]$。利用离散时间傅里叶变换得到功率谱密度：

$$S_{XX}(\mathrm{j}\omega)=\sum_{m=-\infty}^{\infty}R_{XX}[m]\exp(-\mathrm{j}m\omega)=\sigma_X^2,|\omega|\leqslant\pi \tag{8-198}$$

上式定义在 z 平面的单位圆上。平均功率(方差)是给定的功率谱密度下的面积

$$R_{XX}[0]=\frac{1}{2\pi}\int_{-\pi}^{\pi}S_{XX}(\mathrm{j}\omega)\mathrm{d}\omega=\sigma_X^2 \tag{8-199}$$

其中因子 $1/2\pi$ 是由于离散时间傅里叶反变换(见附录 C 中离散时间傅里叶变换的面积性质)。由于功率谱密度在随机序列的有限时间进行积分，方差不是无限的。总结前面的结论，得出下列定义。

定义(白噪声序列) 如果广义平稳随机序列 $X[k]$ 具有平坦的功率谱密度

$$S_{XX}(\mathrm{j}\omega)=\sigma_X^2 \tag{8-200}$$

其中 $\omega\in[-\pi,\pi]$，那么该序列称为**白噪声序列**。它的均值为 0，自相关函数为

$$R_{XX}[m]=\sigma_X^2\delta[m] \tag{8-201}$$

其中 $\sigma_X^2=C_{XX}[0]$ 是随机序列的方差，$C_{XX}[m]=R_{XX}[m]$

假设 $X[k]$ 通过对连续时间过程 $X(t)$ 抽样得到。在这个抽样框架下(第 1 章确定性信号中已经讨论过)，单位圆上的频率范围是 $[-\omega_s/2,\omega_s/2]$，其中 $\omega_s=2\pi f_s$，f_s 是抽样频率。在数字信号模型中通常使用 $[-\pi,\pi]$ 的原因是出于数学便利常取 $f_s=1/T_s=1$，从而 $\pm\omega_s/2=\pm\pi f_s\to\pm\pi$。当仅分析离散时间信号时，这种简化很有用。但是，当需要实现实时(连续)系统时，必须考虑用于产生序列的抽样率，并且有必要使用 $[-\omega_s/2,\omega_s/2]$ 作为频域范围。

第 1 章说明了可以通过变换 $z=\exp(s)$ 从拉普拉斯变换得到 z 变换，该变换将 s 平面上的虚轴映射成 z 平面的单位圆。因此，$[-\omega_s/2,\omega_s/2]$ 上的功率谱密度在 ω_s 的整数倍上重复。

$$S_{X_sX_s}(\omega)=\sigma_X^2\sum_{k=-\infty}^{\infty}\mathrm{rect}(\omega/\omega_s-k)=\sigma_X^2 \tag{8-202}$$

其中不重叠的相邻矩形函数结合在一起，得到图 8-17 所示的所有 $\omega\in\mathcal{R}$ 的平坦响应。功率谱密度也可以写成普通频率 f 和抽样频率 f_s 的形式：

$$S_{X_sX_s}(f)=\sigma_X^2\sum_{k=-\infty}^{\infty}\mathrm{rect}(f/f_s-k)=\sigma_X^2 \tag{8-203}$$

注意 $S_{X_sX_s}(f)$ 是抽样过程 $X_s(t)$ 而不是原来连续过程 $X(t)$ 的功率谱密度。为了对 $X(t)$ 进行抽样并从样本中重建它，$S_{XX}(f)$ 必须是带限信号。从而，$X(t)$ 的频谱实际上是

$$S_{XX}(f)=\sigma_X^2\mathrm{rect}(f/f_s) \tag{8-204}$$

假设 $X(t)$ 的抽样频率是 f_s，如上所述功率谱密度在 f_s 的整数倍上重复，得到平坦的频谱，如图 8-17b 所示。该重复频谱平坦只是因为 $X(t)$ 的频谱是在 $[-f_s/2,f_s/2]$ 上非 0 的矩形函数。(当然，这种理想函数不可能在实际中实现。在这里用它来说明 $X(t)$ 和 $X[k]$ 的联系。)因此，连续时间过程 $X(t)$ 的自相关函数不是 δ 函数，而是

$$R_{XX}(\tau)=\sigma_X^2T_s\mathrm{sinc}(\tau/T_s) \tag{8-205}$$

如图 8-18 所示。白噪声序列无法通过对白噪声过程抽样得到(前面提到的思想在实际中无法实现)。

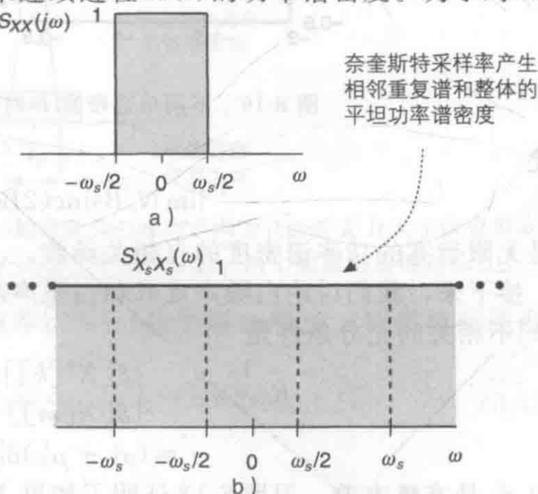


图 8-17 白噪声随机序列 $X[k]$ 的功率谱密度。a) 单位圆上的 $S_{XX}(\mathrm{j}\omega)$；b) 抽样过程 $X_s(t)$ 的 $S_{X_sX_s}(\omega)$

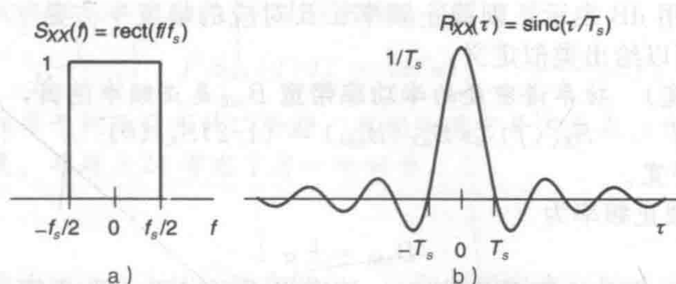


图 8-18　抽样随机过程 $X(t)$。a) 对应白噪声序列 $X[k]$的 $X(t)$的功率谱密度；b) $X(t)$的自相关函数

似乎白噪声过程和白噪声序列之间并不一致，但这是因为白噪声序列定义为离散时刻，仅要求 $T_s=1/f_s$ 的整数倍上的样本不相关。这导致自相关函数为克罗内克 δ 函数。白噪声过程定义在连续时间，要求对于任意 $\tau\neq0$(包括无穷小)，时刻 t 和 $t+\tau$ 上的过程不相关。从而得到自相关函数为狄拉克 δ 函数。

虽然我们不能实现一个白噪声过程，但是它对第 1 章中讨论的连续时间系统的理论模型非常有用。我们可以将随机序列看成通过对图 8-18 所示的具有矩形功率谱密度的非白随机过程抽样得到。该序列只在 T_s 的整数倍上有样本，这正是图 8-18 中非白随机过程 $X(t)$的自相关函数为 0 的时刻。因此，如果矩形覆盖了整个区间$[-f_s/2,f_s/2]$，随机序列的样本不相关。如果过程 $X(t)$的矩形功率谱密度没有覆盖到整个区间，通过抽样得到的随机序列必然相关。例如，如果矩形仅在区间$[-f_s/4,f_s/4]$，那么从傅里叶变换性质得出自相关函数将比第一个过零点 $2T_s$ 宽。从而在 $\tau=T_s$ 时，$R_{XX}(\tau)$非零，再加上 $X[k]$通过抽样频率 $f_s=1/T_s$ 得到，它不能是不相关序列。

前面从理论上说明了白噪声序列如何由带限随机过程抽样得到。在实际中，很容易通过伪随机数发生器，如 MATLAB 中的 rand 或 randn 函数，得到有限方差的白噪声序列。

8.11　带宽

在 8.10 节中，我们介绍了经过滤波器后，白噪声的功率谱密度仅在频率范围 $f\in[-B,B]$上非 0，从而成为带限白噪声过程。该理想频谱可由矩形函数表示：

$$S_{XX}(f)=(N_o/2)\mathrm{rect}(f/2B) \tag{8-206}$$

其自相关是式(8-194)的 sinc 函数。显然，并非所有的功率谱密度都有这样的明确定义的带宽。例如，图 8-19 画出了指数自相关函数 $R_{XX}(\tau)=\exp(-\alpha|\tau|)$的功率谱密度。

$$S_{XX}(f)=(1/\pi)\frac{\alpha}{\alpha^2+f^2} \tag{8-207}$$

其中 $\alpha>0$。对于这种类型的功率谱密度，我们可以定义有效带宽。该方法通常用于滤波器设计，假设定义了单边带宽 B(正频率)，那么传递函数需要满足

$$|H(B_{3\mathrm{dB}})|^2=(1/2)|H(0)|^2 \tag{8-208}$$

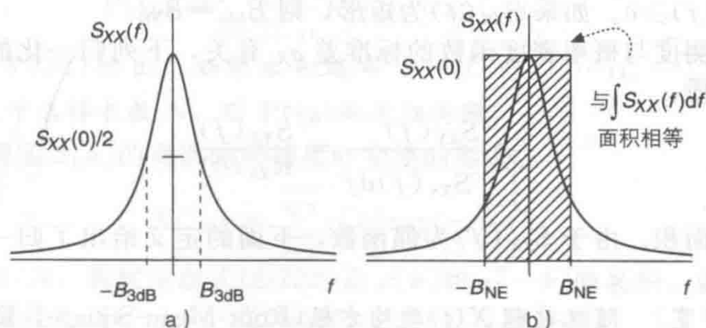


图 8-19　指数自相关函数的功率谱密度；a) 3dB 带宽；b) 噪声等效带宽

如果$|H(f)|^2$用 dB 表示，则截止频率$\pm B$对应的幅度平方是$f=0$最大处的 3dB。对于功率谱密度，可以给出类似定义。

定义（半功率带宽） 功率谱密度的**半功率带宽**$B_{3\text{dB}}$是正频率范围，从而

$$S_{XX}(f)\geqslant S_{XX}(B_{3\text{dB}})=(1/2)S_{XX}(0) \tag{8-209}$$

它也称为 3dB 带宽。

式(8-207)中，截止频率为

$$B_{3\text{dB}}=\pm\alpha \tag{8-210}$$

如图 8-19a 所示，双边 3dB 带宽为2α。该定义也可以用于受载频f_c调制的功率谱密度，例如图 8-15 中的幅度调制信号。对于这种情况，$B_{3\text{dB}}$表达式中的$S_{XX}(0)$需要替换成$S_{XX}(f_c)$。

以下是带宽的另一个测度。

定义（噪声等效带宽） 随机过程$X(t)$的**噪声等效带宽**B_{NE}是

$$B_{\text{NE}}\triangleq(1/2)\frac{\int_{-\infty}^{\infty}S_{XX}(f)\,\mathrm{d}f}{S_{XX}(0)} \tag{8-211}$$

这是带限白噪声与$X(t)$具有相同功率的带宽。

此定义如图 8-19b 所示。基本上，在保证矩形面积与$S_{XX}(f)$相等的前提下，找到高度$S_{XX}(0)$对应的矩形的宽度：

$$\int_{-B_{NE}}^{B_{\text{NE}}}S_{XX}(0)\,\mathrm{d}f=2B_{\text{NE}}S_{XX}(0)=\int_{-\infty}^{\infty}S_{XX}(f)\,\mathrm{d}f \tag{8-212}$$

重新整理该表达式得出定义式。对于指数自相关函数：

$$(1/\pi)\int_{-\infty}^{\infty}\frac{\alpha}{\alpha^2+f^2}\mathrm{d}f=(1/\pi)\arctan(f/\alpha)\Big|_{-\infty}^{\infty}=(1/\pi)(\pi/2+\pi/2)=1 \tag{8-213}$$

其结果与α独立，从而

$$B_{\text{NE}}=\alpha\pi/2 \tag{8-214}$$

如果自相关函数是已知的，则B_{NE}无需对功率谱密度积分：

$$B_{\text{NE}}=(1/2)\frac{R_{XX}(0)}{S_{XX}(0)} \tag{8-215}$$

观察指数自相关函数发现，噪声等效带宽大于 3dB 带宽：$\alpha\pi/2>\alpha$。该结论一般成立：

$$B_{\text{NE}}\geqslant B_{3\text{dB}} \tag{8-216}$$

下面我们给出证明。因为$S_{XX}(f)$为偶函数，仅考虑$f\geqslant0$。另外，因为两种类型的带宽的定义都与$S_{XX}(f)$的最大值有关，不失一般性，令$S_{XX}(0)=1$。然后，

$$\begin{aligned}B_{\text{NE}}&=\int_0^{\infty}S_{XX}(f)\,\mathrm{d}f\\&=\int_0^{B_{3\text{dB}}}S_{XX}(f)\,\mathrm{d}f+\int_{B_{3\text{dB}}}^{\infty}S_{XX}(f)\,\mathrm{d}f\geqslant\int_0^{B_{3\text{dB}}}S_{XX}(f)\,\mathrm{d}f\end{aligned} \tag{8-217}$$

这里，利用了$S_{XX}(f)\geqslant0$。如果$S_{XX}(f)$为矩形，则$B_{3\text{dB}}=B_{\text{NE}}$

带宽的第三个测度与概率密度函数的标准差σ_X有关，下列归一化的功率谱密度具有概率密度函数的性质：

$$\frac{S_{XX}(f)}{\int_{-\infty}^{\infty}S_{XX}(f)\,\mathrm{d}f}=\frac{S_{XX}(f)}{R_{XX}(0)} \tag{8-218}$$

它非负且具有单位面积。由于$S_{XX}(f)$为偶函数，下面的定义给出了归一化功率谱密度的一种类型的标准差。

定义（均方根带宽） 随机过程$X(t)$的均方根(Root-Mean-Square，RMS)带宽为

$$B_{\text{RMS}}\triangleq\sqrt{\int_{-\infty}^{\infty}f^2[S_{XX}(f)/R_{XX}(0)]\,\mathrm{d}f} \tag{8-219}$$

对于前面的指数自相关函数和式(8-213)中的功率谱密度：

$$\frac{1}{R_{XX}(0)}\int_{-\infty}^{\infty} f^2 S_{XX}(f)\mathrm{d}f=(\alpha/\pi)\int_{-\infty}^{\infty}\frac{f^2}{\alpha^2+f^2}\mathrm{d}f \tag{8-220}$$

遗憾的是，这相当于柯西分布的二阶矩，我们知道它并不存在。均方根带宽并不适用于该特定功率谱密度。习题8-24考虑了另一个例子。

8.12　谱估计

随机过程的观测值可以用于估计潜在结构，以期望预测未来的结果。目前没有获得满意长期模型的例子有股市回报和天气预报。在本章的剩余部分，我们将讨论随机过程的功率谱在频率范围内如何分布的信号模型。基本上有两种谱估计方法：(ⅰ)参数法及(ⅱ)非参数法。参数法假定随机过程可以利用输入为白噪声的滤波器表示。非参数法对随机过程的结构不做任何假设。估计自相关函数，然后计算其傅里叶变换，从而导出功率谱密度。这种估计方法称为周期图。

8.12.1　周期图

对于方差有限的广义平稳随机序列 $X[k]$，自相关函数为

$$R_{XX}[m]=\mathcal{E}[X[k]X[k+m]] \tag{8-221}$$

其中，$m\in\mathcal{Z}$是滞后的时移量。考虑 $R_{XX}[m]$的估计：

$$\hat{R}_{XX}[m]=\frac{1}{N}\sum_{k=0}^{N-|m|-1}x[k]x[k+m] \tag{8-222}$$

其中，$x[k]$是随机序列的一个现实，求和中使用$|m|$是因为$R_{XX}[m]$是偶函数。这是单个现实的时间平均，为了使估计有意义，我们假定随机序列具有相关函数各态历经性。由于$|m|\leqslant N-1$，估计中的$|m|$越大，样本越少，N需要足够大才能保证$\hat{R}_{XX}[m]$"捕捉"到相关函数的重要细节。该性质也使得 $m\neq 0$ 时，估计是有偏的：

$$\mathcal{E}[\hat{R}_{XX}[m]]=\frac{1}{N}\sum_{k=0}^{N-|m|-1}R_{XX}[m]=\frac{N-|m|}{N}R_{XX}[m] \tag{8-223}$$

尽管 $N\to\infty$时，它是渐近无偏的。如果需要无偏估计，估计可以进行下列缩放：

$$\frac{N}{N-|m|}\hat{R}_{XX}[m] \tag{8-224}$$

但可以看出，它的方差比式(8-222)中有偏估计的方差大得多。

功率谱密度 $S_{XX}(\mathrm{j}\omega)$是自相关函数 $R_{XX}[m]$的离散时间傅里叶变换。周期图同样也是傅里叶变换，但原函数是自相关函数的有偏估计。

定义(周期图)　随机序列 $X[k]$的**周期图**是自相关函数有偏估计的离散时间傅里叶变换：

$$P(\mathrm{j}\omega)\triangleq\sum_{m=-(N-1)}^{N-1}\hat{R}_{XX}[m]\exp(-\mathrm{j}\omega m) \tag{8-225}$$

其中$\hat{R}_{XX}[m]$由式(8-222)给出。如果求和遍布$\{-(L-1),\cdots,L-1\}$，其中$L<N$但$\hat{R}_{XX}[m]$仍然只依赖于总样本数N，则$P(\mathrm{j}\omega)$称为相关图。

将 $P(\mathrm{j}\omega)$写成现实 $x[k]$的离散时间傅里叶变换的形式：

$$X(\mathrm{j}\omega)=\sum_{k=0}^{N-1}x[k]\exp(-\mathrm{j}\omega k) \tag{8-226}$$

去掉比例因子 $1/N$，观察发现式(8-222)是 $x[k]$和 $x[-k]$的卷积。由于反向(实)序列的离散时间傅里叶变换为 $X(-\mathrm{j}\omega)$，式(8-225)变成(见习题8-25)

$$P(\mathrm{j}\omega)=\frac{1}{N}|X(\mathrm{j}\omega)|^2 \tag{8-227}$$

例 8-11 我们提出了下列随机序列的计算机仿真：

$$X[k]=\cos(2\pi f_1 k+\Phi_1)+2\cos(2\pi f_2 k+\Phi_2)+\cos(2\pi f_3 k+\Phi_3)+V[k] \quad (8\text{-}228)$$

其中 $\{f_1,f_2,f_3\}=\{100,200,300\}$Hz。相位$\{\Phi_n\}$是独立随机变量，在$[-\pi,\pi]$均匀分布，$V[k]$是方差 $\sigma_V^2=0.01$ 的白噪声序列。图 8-20 画出了随机序列的一个现实，以及利用现实的矩形"窗"得到的相应周期图。数据加窗是将现实与某函数逐点相乘，通常它们的持续时间相同；加窗将稍后讨论。从周期图可以清楚看出三个正弦曲线。

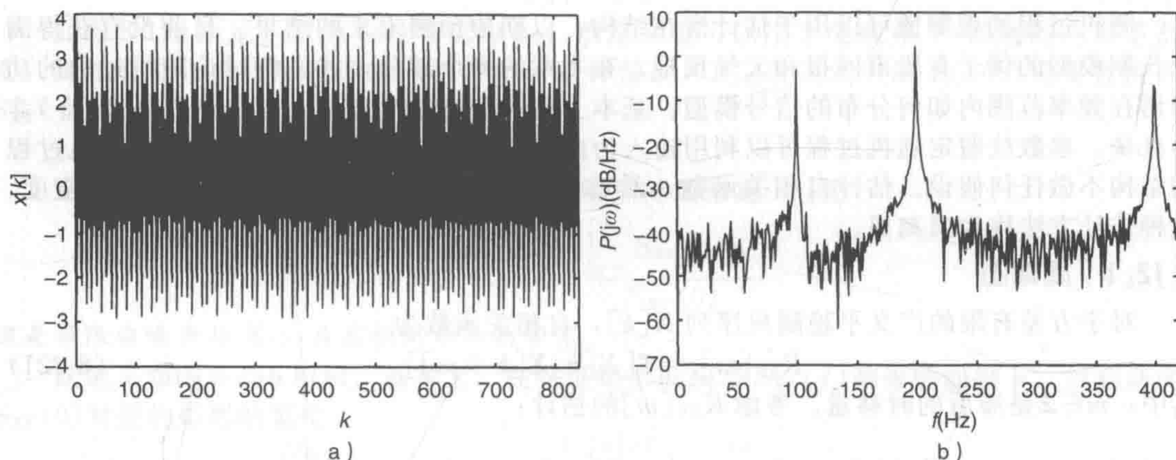


图 8-20 例 8-11 的仿真结果。a) 白噪声三个正弦的现实；b) 利用样本的矩形加窗得到周期图 ◀

接下来，我们推导周期图的均值和方差，以便评价估计的质量。求式(8-225)的期望

$$\mathcal{E}[P(j\omega)]=\sum_{m=-(N-1)}^{N-1}\mathcal{E}[\hat{R}_{XX}[m]]\exp(-j\omega m)$$

$$=\sum_{m=-(N-1)}^{N-1}\frac{N-|m|}{N}R_{XX}[m]\exp(-j\omega m) \quad (8\text{-}229)$$

它是实际的自相关函数乘以三角形函数(窗)$1-|m|/N$ 后的离散时间傅里叶变换。由于加权和离散时间傅里叶变换的极限有限，所以变换只利用了现实的一小部分数据，周期图有偏。在例 8-11 中这种影响很明显，原理上由于噪声理想频谱具有平坦部分，但图 8-20b所示的仿真结果并不是这样。三角加权称为 Bartlett 窗口，式(8-229)的乘积是频域卷积：

$$\mathcal{E}[P(j\omega)]=\frac{1}{2\pi}\int_{-\pi}^{\pi}S_{XX}(j\alpha)W_B(j(\omega-\alpha))d\alpha \quad (8\text{-}230)$$

其中 $S_{XX}(j\omega)$是随机序列功率谱密度

$$W_B(j\omega)=\sum_{m=-(N-1)}^{N-1}w_B[m]\exp(-j\omega m)=(1/N)\frac{\sin^2(\omega N/2)}{\sin^2(\omega/2)} \quad (8\text{-}231)$$

是 Bartlett 窗口的离散时间傅里叶变换：

$$w_B[m]=\begin{cases}1-|m|/N, & |m|\leqslant N-1\\ 0, & \text{其他}\end{cases} \quad (8\text{-}232)$$

图 8-21 画出了 Bartlett 窗口和其离散时间傅里叶变换，其中 $N=15$(式(8-231)中的比例因子 $1/N$ 导致主瓣的最大值为 0dB)。

为了计算周期图的方差，首先计算其自相关函数，它是期望在下列两个不同频率的评价：

$$\mathcal{E}[P(j\omega_1)P(j\omega_2)]=\frac{1}{N^2}\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\sum_{k_3=0}^{N-1}\sum_{k_4=0}^{N-1}\mathcal{E}[X[k_1]X[k_2]X[k_3]X[k_4]]$$

$$\cdot\exp(-j\omega_1(k_1-k_2)-j\omega_2(k_3-k_4)) \quad (8\text{-}233)$$

由于该式难以评价一般的随机序列，我们假定 $X[k]$是高斯白噪声，方差 σ_X^2 为有限值。从而当自变量对相等时：(ⅰ)$k_1=k_2$ 且 $k_3=k_4$，(ⅱ)$k_1=k_3$ 且 $k_2=k_4$(ⅲ)$k_1=k_4$ 且

$k_2=k_3$；可以用 σ_X^4 代替四阶矩（见第 5 章的推导）；否则，期望是零。对于（ⅰ），指数降低到 1，四个求和变成

$$\sum_{k_1=0}^{N-1}\sum_{k_3=0}^{N-1}\sigma_X^4=\sigma_X^4 N^2 \tag{8-234}$$

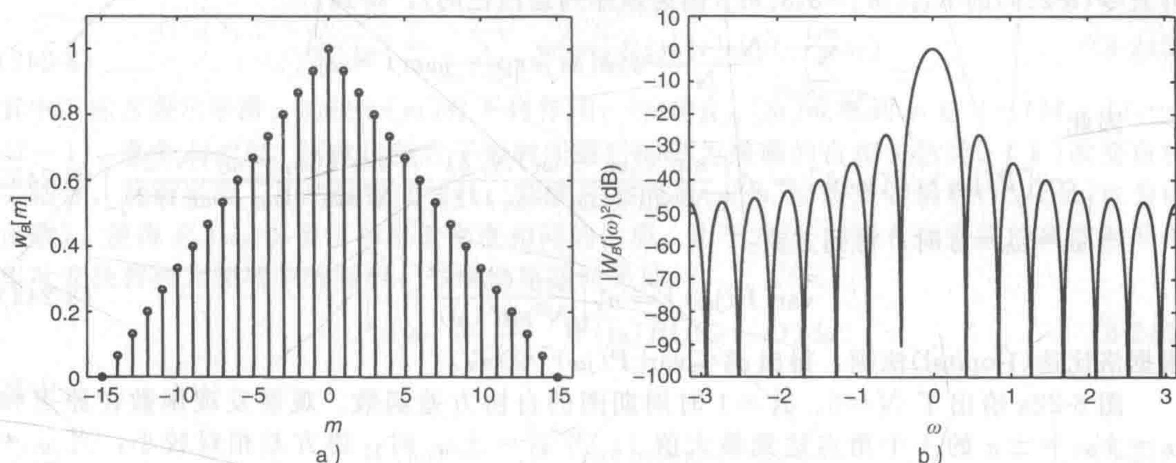


图 8-21　Bartlett 窗的时间域频谱（乘以比例因子 $1/N$），其中 $N=15$。a) Bartlett 窗；b) Bartlett 窗的离散时间傅里叶变换

对于（ⅱ）

$$\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\sigma_X^4\exp(-\mathrm{j}\omega_1(k_1-k_2)-\mathrm{j}\omega_2(k_1-k_2))$$

$$=\sigma_X^4\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\exp(-\mathrm{j}(\omega_1+\omega_2)(k_1-k_2)) \tag{8-235}$$

类似的，对于（ⅲ）

$$\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\sigma_X^4\exp(-\mathrm{j}\omega_1(k_1-k_2)-\mathrm{j}\omega_2(k_2-k_1))=\sigma_X^4\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\exp(-\mathrm{j}(\omega_1-\omega_2)(k_1-k_2)) \tag{8-236}$$

结合这三个结果，得到：

$$\mathcal{E}[P(\mathrm{j}\omega_1)P(\mathrm{j}\omega_2)]=\frac{\sigma_X^4}{N^2}\Big[N^2+\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\exp(-\mathrm{j}(\omega_1+\omega_2)(k_1-k_2))$$

$$+\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\exp(-\mathrm{j}(\omega_1-\omega_2)(k_1-k_2))\Big] \tag{8-237}$$

第一项二重求和，可以看成两个求和的乘积：

$$\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1}\exp(-\mathrm{j}(\omega_1+\omega_2)(k_1-k_2))=\sum_{k_1=0}^{N-1}\exp(-\mathrm{j}(\omega_1+\omega_2)k_1)\sum_{k_2=0}^{N-1}\exp(\mathrm{j}(\omega_1+\omega_2)k_2) \tag{8-238}$$

每一个求和都是在 $\{0,\cdots,N-1\}$ 内的非 0 的矩形序列离散时间傅里叶变换。从附录 C 离散时间傅里叶变换的性质，

$$\sum_{k_1=0}^{N-1}\exp(-\mathrm{j}(\omega_1+\omega_2)k_1)=\exp(-\mathrm{j}(\omega_1+\omega_2)(N-1)/2)\,\frac{\sin((\omega_1+\omega_2)N/2)}{\sin((\omega_1+\omega_2)/2)} \tag{8-239}$$

由于式(8-238)的两个求和构成复共轭对，消去了式(8-239)的相位成分。类似可以得到式(8-237)另一个二重求和的结果，使得自相关函数为

$$\mathcal{E}[P(\mathrm{j}\omega_1)P(\mathrm{j}\omega_2)]=\sigma_X^4\left[1+\frac{\sin^2((\omega_1+\omega_2)N/2)}{N^2\sin^2((\omega_1+\omega_2)/2)}+\frac{\sin^2((\omega_1-\omega_2)N/2)}{N^2\sin^2((\omega_1-\omega_2)/2)}\right]\tag{8-240}$$

自协方差函数为：

$$\operatorname{cov}[P(\mathrm{j}\omega_1),P(\mathrm{j}\omega_2)]=\mathcal{E}[P(\mathrm{j}\omega_1)P(\mathrm{j}\omega_2)]-\mathcal{E}[P(\mathrm{j}\omega_1)]\mathcal{E}[P(\mathrm{j}\omega_2)]\tag{8-241}$$

并且令(8-229)的 $R_{XX}[m]=\sigma_X^2\delta[m]$(因为该序列是白色的)，得到

$$\sum_{m=-(N-1)}^{N-1}\frac{N-|m|}{N}\sigma_X^2\delta[m]\exp(-\mathrm{j}\omega m)=\sigma_X^2\tag{8-242}$$

因此

$$\operatorname{cov}[P(\mathrm{j}\omega_1),P(\mathrm{j}\omega_2)]=\sigma_X^4\left[\frac{\sin^2((\omega_1+\omega_2)N/2)}{N^2\sin^2((\omega_1+\omega_2)/2)}+\frac{\sin^2((\omega_1-\omega_2)N/2)}{N^2\sin^2((\omega_1-\omega_2)/2)}\right]\tag{8-243}$$

当 $\omega_1=\omega_2=\omega$ 时，得到方差：

$$\operatorname{var}[P(\mathrm{j}\omega)]=\sigma_X^4\left[\frac{\sin^2(\omega N)}{N^2\sin^2(\omega)}+1\right]\tag{8-244}$$

根据洛比达(Ĥopital)法则，得出 $\sigma_X^4\leqslant\operatorname{var}[P(\mathrm{j}\omega)]\leqslant 2\sigma_X^4$。

图 8-22a 给出了 $N=5$、$\sigma_X^2=1$ 时周期图的自协方差函数。观察发现函数在原点和 $\omega_1=\pm\omega_2=\pm\pi$ 的 4 个角点达到最大值 1。当 $\omega_1=\pm\omega_2$ 时，协方差相对较小，当 $\omega_1=2\pi m/N$ 且 $\omega_2=2\pi n/N$ 时，其中 m，$n\in\mathcal{Z}$且 $m\neq n$ 时，恰恰是零。图 8-22b 所示的方差相当于在 $\omega=\omega_1=\omega_2$时对函数进行"切片"。当 $\omega_1\neq\pm\omega_2$ 时，随着 N 的增加，协方差减小，在原点和 $\omega_1=\pm\omega_2=\pm\pi$ 的 4 个角点处，值仍然是 2，当 $\omega_1=\pm\omega_2$ 时，减小到 1。这表现在图 8-22c 所示 $N=33$ 时的方差中。随着 N 的增加，方差并不会减小到 0(通常我们选择方差持续减小的估计量，第 9 章将讨论)。

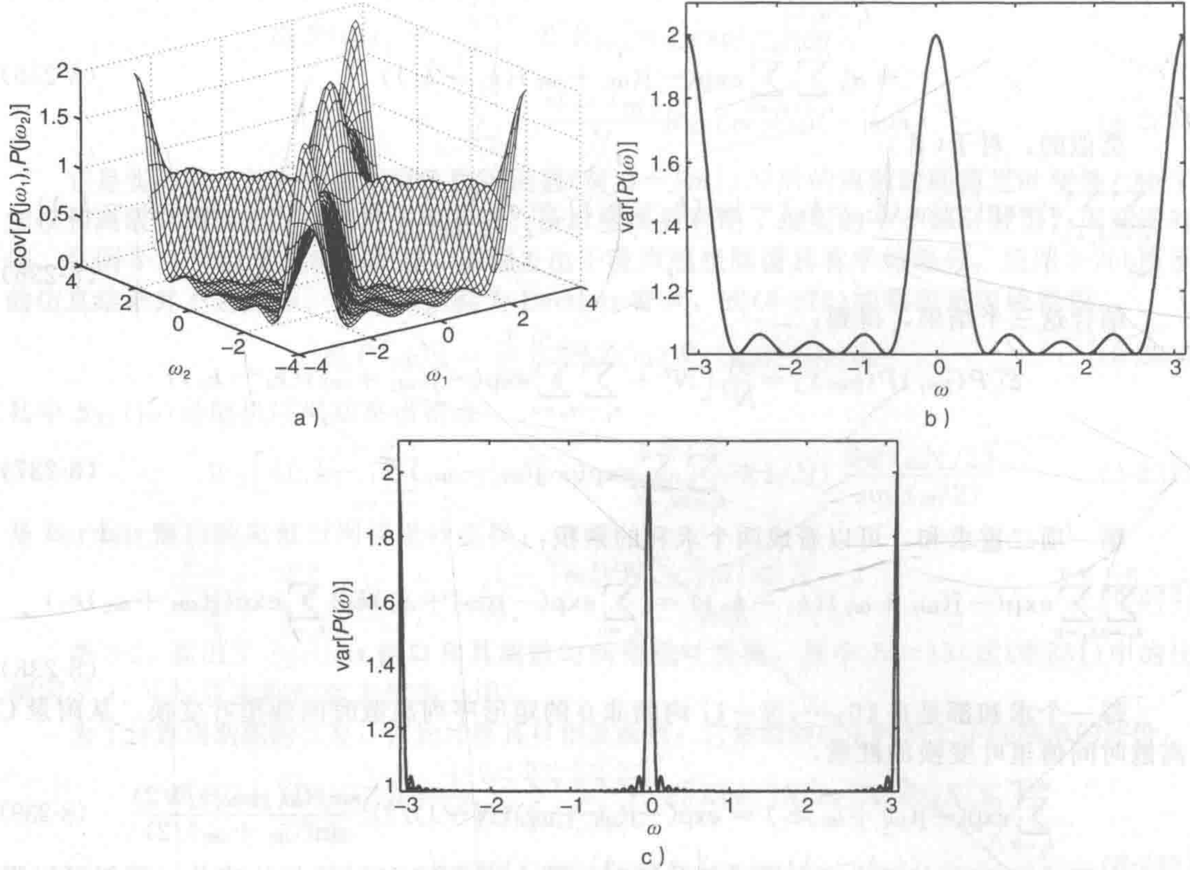


图 8-22　方差 $\sigma_X^2=1$ 的高斯白噪声序列的自协方差函数和周期图。a) 周期图的自协方差函数 $\operatorname{cov}[P(\mathrm{j}\omega_1),P(\mathrm{j}\omega_2)]$，$N=5$；b) 周期图的方差 $\operatorname{var}[P(\mathrm{j}\omega)]$，$N=5$；c) 周期图的方差 $\operatorname{var}[P(\mathrm{j}\omega)]$，$N=33$

在下面的章节中，我们利用两个标准周期图的变量估计随机序列的功率谱密度：(ⅰ)平滑周期图；(ⅱ)现实加窗。

8.12.2 平滑周期图

周期图的一个变化是对$\hat{R}_{XX}[m]$加窗，也就是让它乘以一个函数：

$$P_S(j\omega)=\sum_{m=-(N-1)}^{N-1}w[m]\hat{R}_{XX}[m]\exp(-j\omega m) \tag{8-245}$$

其中下标S表示平滑。加权$w[m]$有下列作用：(ⅰ)将$\hat{R}_{XX}[m]$截断到$m\in\{-(M-1),\cdots,M-1\}$，通常$M\ll N$，因此仅包含了短时间滞后的更为准确的自相关估计。(ⅱ)改变自相关估计，从而平滑了功率谱密度估计。需要注意的是$w[m]$应该是偶函数($R_{XX}[m]$也为偶函数)，使得$P_S(j\omega)$保留了功率谱密度相同的性质。由于式(8-245)右边乘积的离散时间傅里叶变换替换为频域中的卷积，平滑性质显而易见。

$$P_S(j\omega)=\frac{1}{2\pi}\int_{-\pi}^{\pi}W(j\alpha)P(j(\omega-\alpha))d\alpha \tag{8-246}$$

其中

$$W(j\omega)=\sum_{m=-(M-1)}^{M-1}w[m]\exp(-j\omega m) \tag{8-247}$$

是窗口的离散时间傅里叶变换，$P(j\omega)$是标准的周期图。由于$M\ll N$，对应于窄窗口，其傅里叶变换$W(j\omega)$相对较宽，因此它与周期图的卷积使得$P_S(j\omega)$比$P(j\omega)$更平滑。

对于标准周期图$P(j\omega)$，我们发现$\varepsilon[P(j\omega)]$为$R_{XX}[m]$的加权形式的离散时间傅里叶变换，其中权值是总长度为$2N-1$的Bartlett窗口。由于$\hat{R}_{XX}[m]$是$R_{XX}[m]$的有偏估计，我们用三角加权来得到比无偏估计更小的方差。此处描述的用于平滑周期图加窗的不同之处在于它是有意用于自相关函数的估计$\hat{R}_{XX}[m]$。此外，由于$M\ll N$，长时间滞后的$\hat{R}_{XX}[m]$将被删除；如前面提到的，受较小的求和上限$N-|m|-1$的限制，那些自相关估计不太准确。

$P_S(j\omega)$的均值为：

$$\mathcal{E}[P_S(j\omega)]=\sum_{m=-(N-1)}^{N-1}w[m]w_B[m]R_{XX}[m]\exp(-j\omega m) \tag{8-248}$$

这里包含了出现在式(8-229)中的Bartlett窗口$w_B[m]$。因此，实际的自相关函数$R_{XX}[m]$由复合窗$w[m]w_B[m]$加权。然而，由于$M\ll N$，在$\{-(M-1),\cdots,M-1\}$上，Bartlett窗相对平坦，使得$w[m]$主导窗口乘积$w[m]w_B[m]\approx w[m]$。$w_B[m]$在$\{-(M-1),\cdots,M-1\}$上的锥度对均值影响不大。在频域，我们有以下近似

$$\mathcal{E}[P_S(j\omega)]\approx\frac{1}{2\pi}\int_{-\pi}^{\pi}W(j\alpha)S_{XX}(j(\omega-\alpha))d\alpha \tag{8-249}$$

其中，$W(j\alpha)$是明确的用于自相关估计加窗的离散时间傅里叶变换。

平滑周期图的自相关函数为：

$$\begin{aligned}\operatorname{cov}[P_S(j\omega_1),P_S(j\omega_2)]=&\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}W(j\alpha_1)W(j\alpha_2)\\&\cdot\operatorname{cov}[P(j(\omega_1-\alpha_1)),P(j(\omega_2-\alpha_2))]d\alpha_1 d\alpha_2\end{aligned} \tag{8-250}$$

类似标准周期图分析时的做法，我们假定$X[k]$是方差为σ_X^2的高斯白噪声，从而积分中的协方差可从式(8-243)得到。为了简化表示式，假设N比较大，使得式(8-231)中Bartlett窗的离散时间傅里叶变换可由δ函数近似(Oppenheim和Schafer，1975；Tretter，1976)：

$$W_B(j\omega)=\frac{\sin^2(\omega N/2)}{N\sin^2(\omega/2)}\approx\sum_{n=-\infty}^{\infty}2\pi\delta(\omega-2\pi n) \tag{8-251}$$

利用洛比达法则很容易得出$W_B(j\omega)$在2π的整数倍上有峰值N。随着N的增加，峰

值趋于无穷大，并且每个“脉冲”$W_B(j\omega)$的宽度趋近于零。由于 $w_B[m]$是偶函数，功率谱密度的面积性质可以写成：

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sin^2(\omega N/2)}{N\sin^2(\omega/2)}d\omega = w_B[0] = 1 \tag{8-252}$$

从图 8-23 的例子中，我们看到随着 N 的增加，Bartlett 窗的离散时间傅里叶变换在 $\omega\in[-\pi,\ \pi]$上趋于面积为 2π 的 δ 函数，并且在 2π 的整数倍频率上重复出现。正是这种窄 Bartlett 离散时间傅里叶变换脉冲，导致了图 8-22c 在原点出现脉冲状的方差。

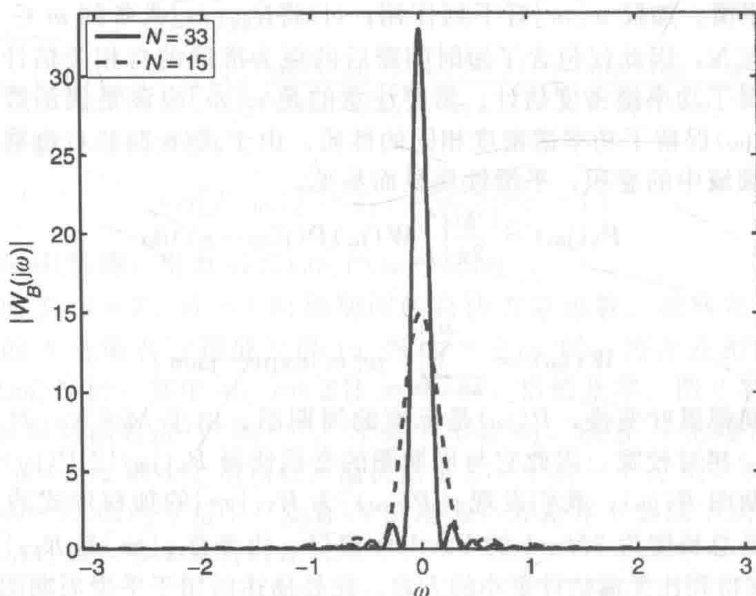


图 8-23 Bartlett 窗的频谱(乘以 $1/N$，线性比例因子)

将式(8-251)的近似代入式(8-243)，在频率范围 $\omega\in[-\pi,\ \pi]$内，得出两个 δ 函数：

$$\sigma_X^4\left[\frac{\sin^2((\omega_1+\omega_2)N/2)}{N^2\sin^2((\omega_1+\omega_2)/2)}+\frac{\sin^2((\omega_1-\omega_2)N/2)}{N^2\sin^2((\omega_1-\omega_2)/2)}\right]\approx\frac{2\pi\sigma_X^4}{N}[\delta(\omega_1+\omega_2)+\delta(\omega_1-\omega_2)] \tag{8-253}$$

使用适当的参数将上式代入式(8-250)，将二重积分变成一重积分：

$$\begin{aligned}\operatorname{cov}[P_S(j\omega_1),P_S(j\omega_2)]\approx\frac{\sigma_X^4}{2\pi N}\int_{-\pi}^{\pi}&[W(j(\omega_1+\omega_2-\alpha_2))\\&+W(j(\omega_1-\omega_2+\alpha_2))]W(j\alpha_2)d\alpha_2\end{aligned} \tag{8-254}$$

当 $\omega_1=\omega_2=\omega$ 时，得到方差：

$$\operatorname{var}[P_S(j\omega)]\approx\frac{\sigma_X^4}{2\pi N}\int_{-\pi}^{\pi}[W(j(2\omega-\alpha_2))W(j\alpha_2)+W^2(j\alpha_2)]d\alpha_2 \tag{8-255}$$

假设 M 足够大(但仍然$\ll N$)，从而 $W(j\omega)$较窄，与第二项相比，第一项积分比较小，可以忽略不计。得到近似

$$\operatorname{var}[P_S(j\omega)]\approx\frac{\sigma_X^4}{2\pi N}\int_{-\pi}^{\pi}W^2(j\alpha_2)d\alpha_2 \tag{8-256}$$

为了与式(8-244)中 $P(j\omega)$的方差进行比较，我们需要得到积分上限。利用帕斯瓦尔定理：

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}W^2(j\alpha_2)d\alpha_2=\sum_{-(M-1)}^{M-1}w^2[m]\leqslant 2M-1 \tag{8-257}$$

其中窗口已经归一化，使得对于任意 m，都有 $0\leqslant w[m]\leqslant 1$，以得到积分上限。

$$\operatorname{var}[P_S(j\omega)]\approx(1/N)(2M-1)\sigma_X^4 \tag{8-258}$$

为了与式(8-244)中 $P(j\omega)$方差的下界 σ_X^4 比较，我们发现当窗口宽度 M 远小于自相关

估计的样本 N 时，$\mathrm{var}[P_S(j\omega)]$减小。而且，通过截断估计的自相关函数，时间滞后较大和方差较大的项被排除在外。然而，当M较小时，窗口的离散时间傅里叶变换比较宽，这意味着该周期图不能为各功率谱密度分量提供好的分辨率，特别是当它们的频率很近时。

例 8-12 考虑使用下列矩形窗，其中 $M<N$：

$$w_R[m]=\begin{cases}1, & |m|\leqslant M-1\\ 0, & \text{其他}\end{cases} \tag{8-259}$$

其离散时间傅里叶变换为

$$W_R(j\omega)=\frac{\sin((2M-1)\omega/2)}{\sin(\omega/2)} \tag{8-260}$$

图 8-24 画出了矩形窗的一个例子，可以看到它的主瓣比 Bartlett 窗窄，但旁瓣较高。众所周知，这是离散时间信号处理在主瓣宽度（通带内）和旁瓣高度（阻带内）之间的权衡。然而，由于 $W_R(j\omega)$存在负值而功率谱密度估计必须非负，所以矩形窗不能用于平滑周期图。而 Bartlett 窗可用，因为从(8-231)明显看出，它的离散时间傅里叶变换非负。图 8-25 画出了例 8-11 中具有三个正弦且 $M=N/2$ 的信号的平滑周期图。观察发现，即使频谱噪声看起来比图 8-20 的小（方差降低），但其分辨率较差。将标准周期图与 Bartlett 窗的频谱卷积导致峰加宽，这样间隔较小的正弦信号可能更难以分开。这个例子说明，即使估计的方差可以减少，当考虑估计的其他指标时，如分辨和分离信号，整体质量可能不会提高。

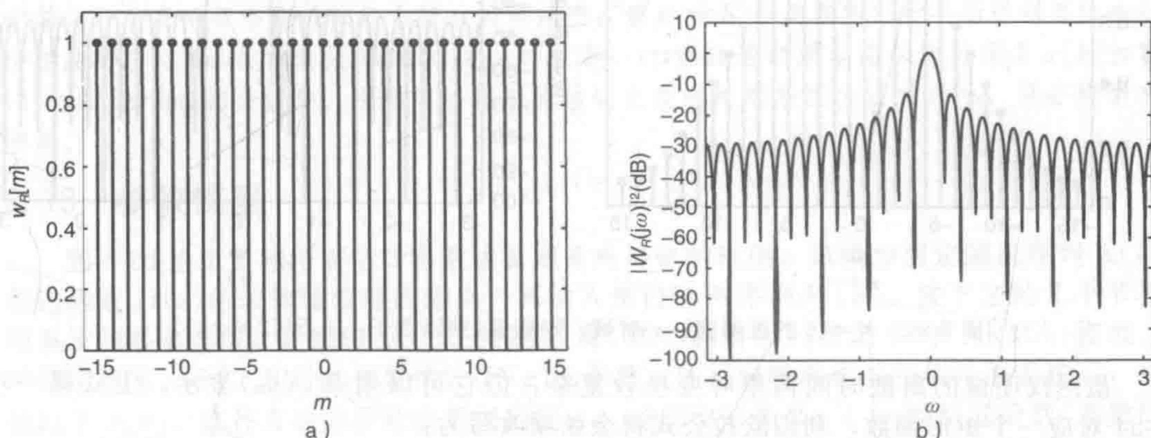


图 8-24　$M=16$ 的矩形窗口。a) 时域；b)频谱（乘以因子 $1/(2M-1)$）

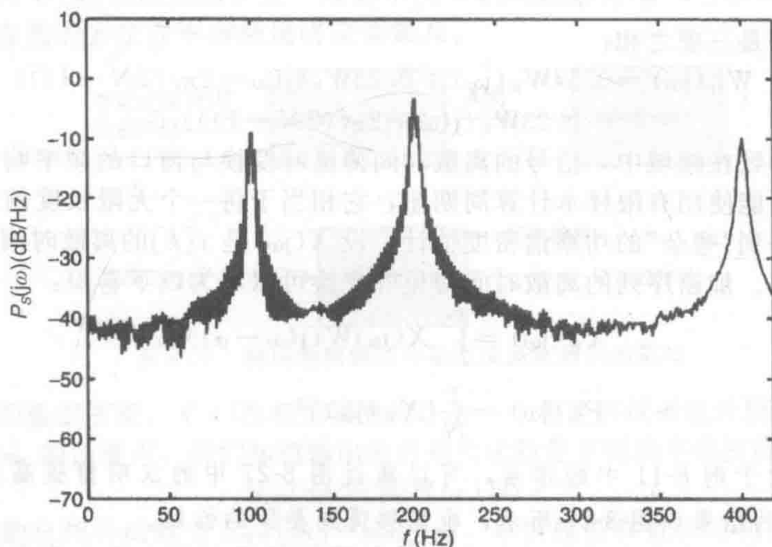


图 8-25　例 8-12 的平滑周期图

8.12.3　改进的周期图

另一种形式的标准周期图是对它的实现本身加窗(而不是自相关估计)，得到改进的周期图。

定义(改进的周期图)　随机序列 $X[k]$ 的**改进周期图**为：

$$P_M(\mathrm{j}\omega) \triangleq \frac{1}{\sum_{k=0}^{N-1} w^2[k]} \left| \sum_{k=0}^{N-1} w[k]x[k]\exp(-\mathrm{j}\omega k) \right|^2 \tag{8-261}$$

其中 $w[k]$ 是现实 $x[k]$ 的窗口。

如果窗口是图 8-24 所示的矩形，当窗口长度为 $2N-1$ 时，改进的周期图变成标准周期图。观察发现，矩形窗频率响应的主瓣宽度较窄，但与图 8-26 所示的汉明窗相比，旁瓣较高。汉明窗是以下升余弦函数：

$$w_{\mathrm{H}}[k] = \begin{cases} 0.54 + 0.46\cos(2\pi k/(2N-1)), & |k| \leqslant N-1 \\ 0, & \text{其他} \end{cases} \tag{8-262}$$

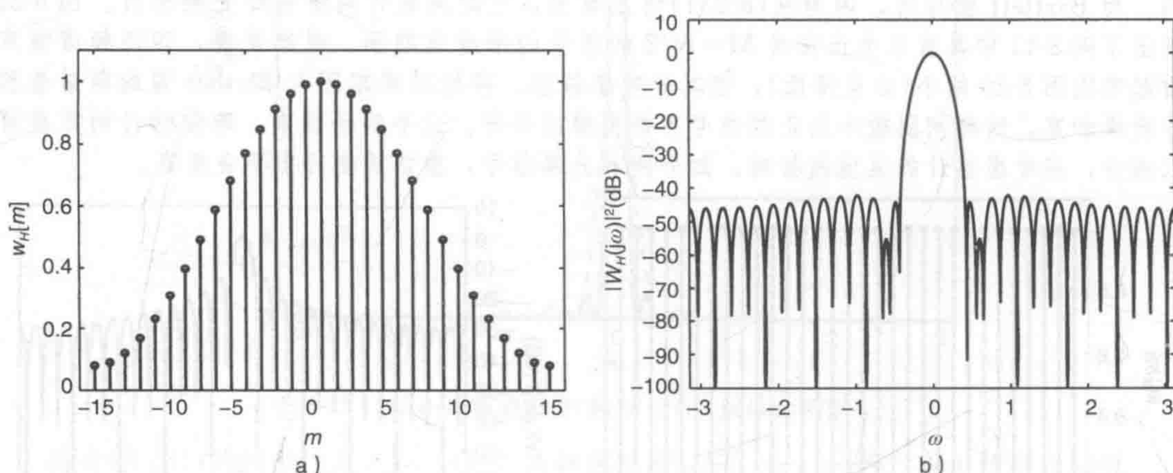


图 8-26　$N=16$ 的汉明窗。a)时域；b)频谱(乘以因子 $1/0.54(2N-1)$)

虽然汉明窗的离散时间傅里叶变换较复杂，但它可以用 $W_{\mathrm{R}}(\mathrm{j}\omega)$ 表示。上式第一项 0.54 对应一个矩形函数，利用欧拉公式将余弦项改写为：

$$0.46\cos(2\pi m/(2N-1)) = 0.23\exp(\mathrm{j}2\pi m/(2N-1)) + 0.23\exp(-\mathrm{j}2\pi m/(2N-1)) \tag{8-263}$$

因此，频谱是三项之和：

$$\begin{aligned} W_{\mathrm{H}}(\mathrm{j}\omega) = {} & 0.54W_{\mathrm{R}}(\mathrm{j}\omega) + 0.23W_{\mathrm{R}}(\mathrm{j}(\omega - 2\pi/(2N-1))) \\ & + 0.23W_{\mathrm{R}}(\mathrm{j}(\omega + 2\pi/(2N-1))) \end{aligned} \tag{8-264}$$

序列加窗导致在频域中，信号的离散时间傅里叶变换与窗口的频率响应相卷积。这就解释了为什么只能使用有限样本计算周期图，它相当于将一个无限长度的序列乘以有限长度的矩形窗，得到“嘈杂”的功率谱密度估计。设 $X(\mathrm{j}\omega)$ 是 $x[k]$ 的离散时间傅里叶变换，假定它的长度无限。加窗序列的离散时间傅里叶变换可以写为以下卷积：

$$X_{\mathrm{W}}(\mathrm{j}\omega) = \int_{-\pi}^{\pi} X(\mathrm{j}\alpha)W(\mathrm{j}(\omega-\alpha))\mathrm{d}\alpha \tag{8-265}$$

$$P_{\mathrm{M}}(\mathrm{j}\omega) = \frac{1}{N}|X_{\mathrm{W}}(\mathrm{j}\omega)|^2 \tag{8-266}$$

例 8-13　对于例 8-11 中的信号，可以通过图 8-27 中的汉明窗提高旁瓣抑制。利用 Bartlett 窗对应的结果如图 8-27 所示，也能够得到类似的性能。

需要注意的是，改进的周期图和平滑周期图的不同之处在于，前者是对随机序列加窗。频域中的相应卷积降低了序列中加性噪声的影响，从而得到更准确的自相关估计。

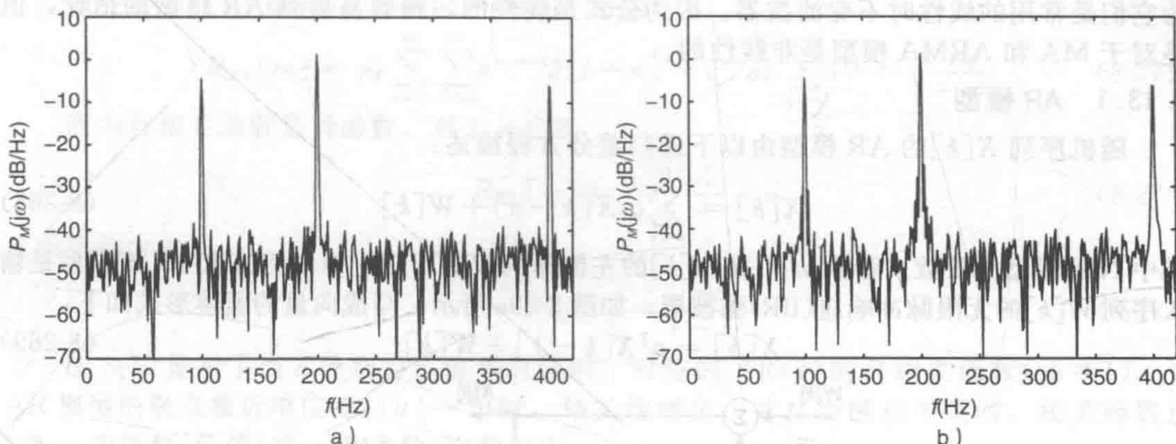


图 8-27　例 8-13 中的改进的周期图使用。a) 汉明窗；b)Bartlett 窗口

平滑周期图不对现实而是对自相关估计加窗。工程中的做法通常是在计算自相关估计前，对原始样本进行处理(加窗或滤波)，而不是对自相关函数进行处理(平滑)。

最后，我们简单介绍两种称为 Bartlett 和 Welch 方法的平均法，这两种方法对标准周期图而不是随机序列的现实进行处理。Bartlett 方法将现实 $x[k]$的 N 个样本分成 K 个互不相重的部分，然后用矩形窗计算每个部分的周期图，最后对 K 个周期图进行平均得到整体的功率谱密度估计。Welch 方法是 Bartlett 方法的扩展：(i)像改进的周期图那样对现实 $x[k]$加窗；(ii)允许 Bartlett 部分重叠。这些方法得到的估计比改进的周期图方法噪声小，但分辨率有时较低。◀

8.13　参数模型

图 8-28 给出了利用谱估计参数法实现*系统辨识*的框图。该模型假定随机序列 $X[k]$是传递函数 $H(z)$的未知滤波器的输出，其输入是白噪声序列 $W[k]$。接下来的几小节将介绍系统的基本类型：自回归(AR)、滑动平均(MA)和自回归滑动平均(ARMA)模型。对于特定现实 $x[k]$，其目的是要估计 $H(z)$的参数，使得白噪声通过所估计的系统后，输出接近于 $x[k]$。这种方法需要解决下列问题：(i)模型的类型；(ii)模型的阶数(参数的个数)；(iii)计算参数值的方法。由于参数估计仅从输出 $X[k]$得到，该建模方法有时也称为“盲”，因为输入 $W[k]$无法观测。更一般的系统辨识问题假定输入与输出可测，这种情况下估计滤波器参数的方法在本章结尾将简要提及。

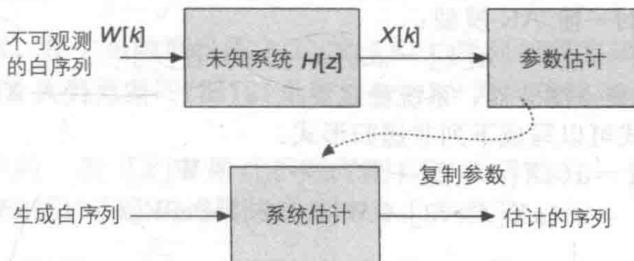


图 8-28　利用参数频谱分析实现系统辨识的框架

鉴于前面的模型框架，$X[k]$的相关函数将作为一个重要特征来估计模型参数。由于输入信号是方差为 σ_W^2 的白噪声，我们知道输出的自相关函数是下列功率谱密度的 z 反变换：

$$S_{YY}(z) = \sigma_W^2 H(z) H(z^{-1}) \tag{8-267}$$

由于输入的自相关函数为 $R_{WW}[m]=\sigma_W^2\delta[m]$，显然模型参数决定相关程度，因此模型选择是一个重要的考虑因素。在包含自然信号的应用中，序列实际上比传递函数 $H(z)$表示的更复杂。这些模型仅是随机序列如何产生的近似。考虑 AR、MA 和 ARMA 模型是因

为它们是常用的线性时不变滤波器。因为公式是线性的，很容易得到 AR 模型的估计，但是对于 MA 和 ARMA 模型是非线性的。

8.13.1 AR 模型

随机序列 $X[k]$的 AR 模型由以下递归差分方程描述：

$$X[k]=\sum_{n=1}^{N}a_nX[k-n]+W[k] \tag{8-268}$$

其中 N 是模型的阶数。参数 a_0 可在 $X[k]$的左侧作为权值出现，但常将其取 1。该模型是输入序列 $W[k]$的无限脉冲响应(IIR)滤波器，如图 8-29a 所示。写成向量的紧凑形式如下：

$$X[k]=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}[k-1]+W[k] \tag{8-269}$$

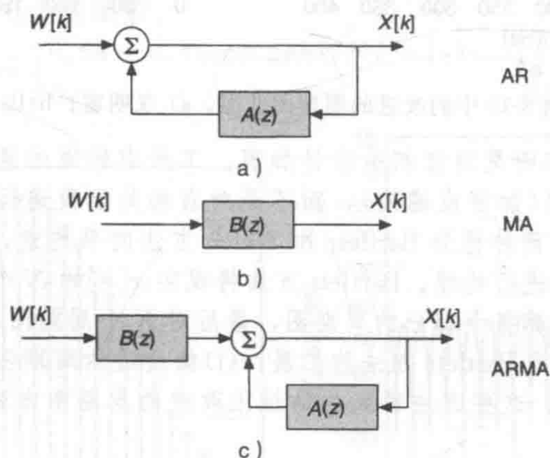


图 8-29 随机序列模型。a) AR；b) MA；c) ARMA

其中，$\boldsymbol{a}\triangleq[a_1,\cdots,a_N]^{\mathrm{T}}$，$\boldsymbol{X}[k-1]\triangleq[X[k-1],\cdots X[k-N]]^{\mathrm{T}}$，系统传递函数是

$$H(z)=\frac{1}{1-\sum_{n=1}^{N}a_nz^{-n}}\triangleq\frac{1}{1-A(z)} \tag{8-270}$$

对于 AR 模型，$H(z)$只有极点；包含极点和零点的更一般的情况是后面讨论的 ARMA 模型。对于因果稳定系统，$H(z)$的极点都在单位圆内。由于 $W[k]$是白序列，$X[k]$的功率谱密度满足

$$S_{XX}(z)=H(z)H(z^{-1})=\frac{\sigma_W^2}{[1-A(z)][1-A(z^{-1})]} \tag{8-271}$$

考虑下列 $N=1$ 的一阶 AR 模型：

$$X[k]=aX[k-1]+W[k] \tag{8-272}$$

这是极点在 $z=a$ 的单极点滤波器，系统稳定要求$|a|<1$。依次代入 $X[k-1]$、$X[k-2]$，等等，上面的递归表达式可以写成下列非递归形式：

$$\begin{aligned}X[k]&=a(aX[k-2]+W[k-1])+W[k]\\&=a^2(aX[k-3]+W[k-2])+aW[k-1]+W[k]\\&\ \ \vdots\\&=\sum_{n=0}^{\infty}a^nW[k-n]\end{aligned} \tag{8-273}$$

由于 $W[k]$具有零均值，对于所有的 k 都有$\mathcal{E}[X[k]]=0$。自相关函数是

$$\begin{aligned}\mathcal{E}[X[k]X[k+m]]&=\sum_{n=0}^{\infty}\sum_{p=0}^{\infty}a^{n+p}\,\mathcal{E}[W[k-p]W[k+m-n]]\\&=\sum_{n=0}^{\infty}\sum_{p=0}^{\infty}a^{n+p}\sigma_W^2\delta[m-n+p]\end{aligned} \tag{8-274}$$

做变量替换 $l=m+p$，并假定 m 为正，得到

$$R_{XX}[m]=\sigma_W^2\sum_{n=0}^{\infty}\sum_{l=m}^{\infty}a^{n-m+l}\delta[l-n]=a^{-m}\sigma_W^2\sum_{n=m}^{\infty}a^{2n}=\frac{a^m\sigma_W^2}{1-a^2}\tag{8-275}$$

因为自相关函数是偶函数，对于 $m\in\mathbb{Z}$：

$$R_{XX}[m]=\frac{a^{|m|}\sigma_W^2}{1-a^2}\tag{8-276}$$

从而得到方差

$$\sigma_X^2=R_{XX}[0]=\frac{\sigma_W^2}{1-a^2}\tag{8-277}$$

图 8-30 显示了当 a 分别为正值和负值时，对应的 AR(1)的自相关函数($\sigma_W^2=1$)。当 AR 模型的极点靠近单位圆($|a|\rightarrow 1$)时，相关性增加，并且当极点为负时，相关函数正(当 m 为偶数)负值(当 m 为奇数)交替出现。

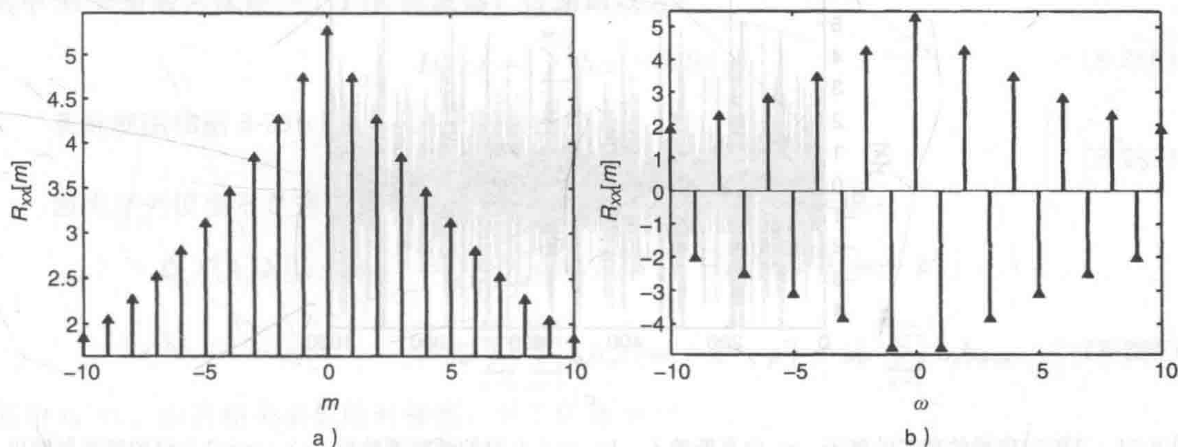


图 8-30 一个 AR(1)序列的自相关函数。a) $a=0.9$；b) $a=-0.9$

自相关函数也可以通过功率谱密度得到：

$$S_{XX}(z)=\frac{\sigma_W^2}{(1-az^{-1})(1-az)}=\frac{\sigma_W^2}{1-a^2}\left(\frac{1}{1-az^{-1}}+\frac{1}{1-az}\right)\tag{8-278}$$

已经进行了部分分式展开(Partical Fraction Expansion，PFE)。第一项对应一个右边函数，第二项对应一个左边函数。z 反变换是

$$R_{XX}[m]=\frac{\sigma_W^2}{1-a^2}(a^{-m}u[m]+a^{m}u[-m+1])=\frac{a^{-|m|}\sigma_W^2}{1-a^2}\tag{8-279}$$

与式(8-276)结果相同。相关函数满足图 8-30 的 AR(1)序列的现实如图 8-31 所示。观察发现输出信号比输入信号具有更多的“结构”，并且 $a=0.9$ 对应的输出看起来比 $a=-0.9$ 更好预测。

对于式(8-268)中的一般 AR 序列，不能得到自相关函数的一个简单表达式。相反，我们可以用一个递归表达式近似描述随机序列本身。考虑

$$\mathcal{E}[X[k-m]X[k]]=\sum_{n=1}^{N}a_n\,\mathcal{E}[X[k-m]X[k-n]]+\mathcal{E}[X[k-m]W[k]]\tag{8-280}$$

其中 $m>0$。等式右边的第二项是零，因为 $X[k]$的以前的样本与白序列 $W[k]$独立。从而得到下列递归表达式：

$$R_{XX}[m]=\sum_{n=1}^{N}a_nR_{XX}[m-n]\tag{8-281}$$

对于不同 m 值，可以写成矩阵形式：

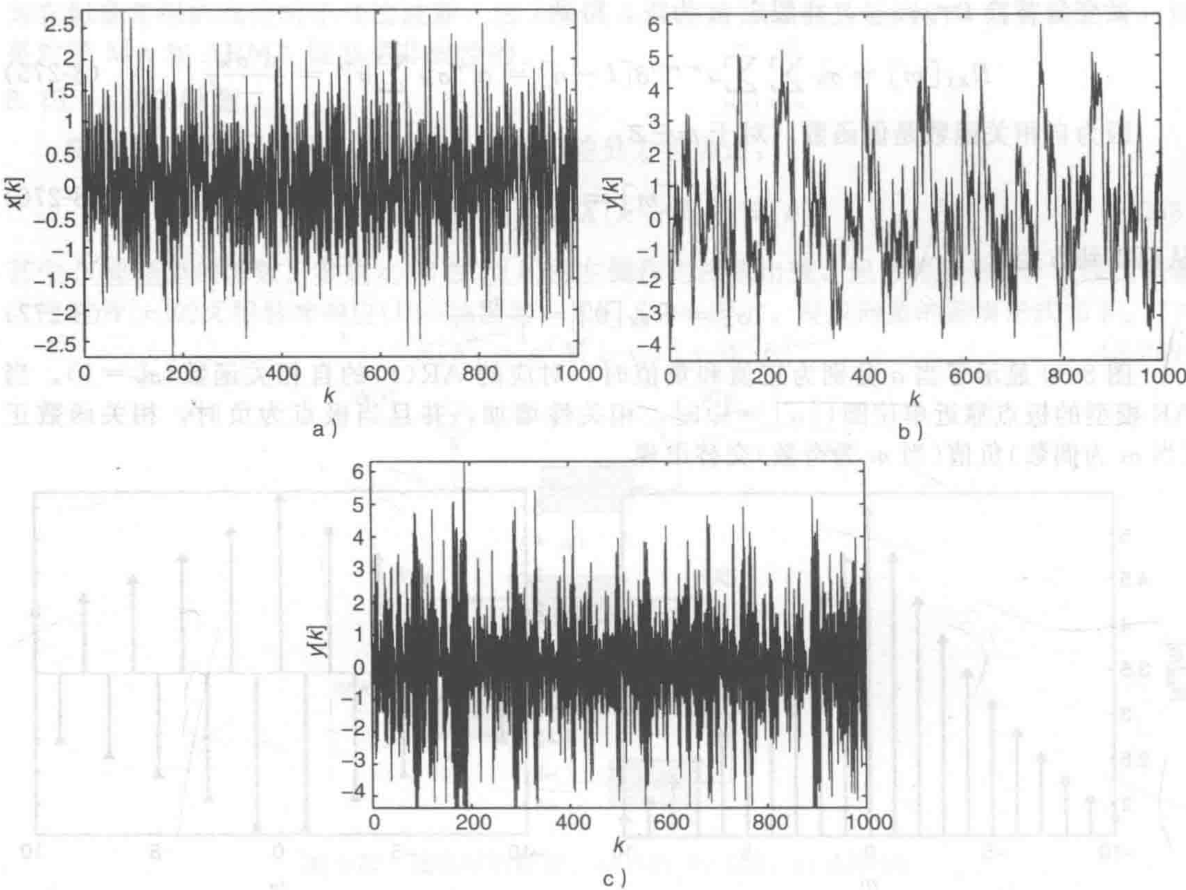


图 8-31 AR(1)序列的现实的例子。a) 白高斯输入；b) $a=0.9$ 时的滤波器输出；c) $a=-0.9$ 时的滤波器输出

$$\begin{bmatrix} R_{XX}[1] \\ R_{XX}[2] \\ \vdots \end{bmatrix} = \begin{bmatrix} R_{XX}[0] & R_{XX}[-1] & \cdots & R_{XX}[N-1] \\ R_{XX}[1] & R_{XX}[0] & \cdots & R_{XX}[N-2] \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \tag{8-282}$$

对于 $m=0$：

$$\mathcal{E}[X^2[k]] = \sum_{n=1}^{N} a_n \mathcal{E}[X[k-n]X[k]] + \mathcal{E}[W[k]X[k]] \tag{8-283}$$

可得：

$$R_{XX}[0] = \sum_{n=1}^{N} a_n R_{XX}[n] + \sigma_W^2 \tag{8-284}$$

由于自相关函数是偶函数，式(8-282)中的参数可以改变，使得所有参数为正，表明式(8-282)中的矩阵是 Toepliez 矩阵。用于求解滤波器系数$\{a_n\}$的方程称为 Yule-Walker 方程，可以将它们代入式(8-284)来估计方差 σ_W^2。当然，由于 $R_{XX}[m]$未知，可以从 $X[k]$的现实得到一个估计。

例 8-14 假设估计的自相关函数恰好有理想的形式$\hat{R}_{XX}[m]=(0.9)^{|m|}$，考虑 Yule-Walker 方程的 $N=3$，从而

$$\begin{bmatrix} 0.9 \\ 0.81 \\ 0.729 \end{bmatrix} = \begin{bmatrix} 1 & 0.9 & 0.81 \\ 0.9 & 1 & 0.9 \\ 0.81 & 0.9 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \tag{8-285}$$

求解得到系数 $[a_1,a_2,a_3]^{\mathrm{T}}=[0.9,0,0]^{\mathrm{T}}$，从式(8-284)得到噪声方差是 $\sigma_W^2=1-0.81=0.19$。得到一阶系统并不奇怪，因为这个特定的自相关函数具有式(8-276)的形式，并且

根据式(8-277)σ_W^2 与 $R_{XX}[0]$有关。假设由于不精确的自相关测量，Yule-Walker 方程为

$$\begin{bmatrix}0.9\\0.8\\0.7\end{bmatrix}=\begin{bmatrix}1 & 0.9 & 0.8\\0.9 & 1 & 0.9\\0.8 & 0.9 & 1\end{bmatrix}\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix} \tag{8-286}$$

求解这些方程得到 $[a_1,a_2,a_3]^{\mathrm{T}}=[0.9444,0,-0.0556]^{\mathrm{T}}$，从式(8-277)得到噪声方差 $\sigma_W^2=1-(0.9444)\times(0.9)+(0.0556)\times(0.7)\approx 0.1905$。估计关于测量误差的灵敏度是一个重要的问题，特别是 AR 模型的参数较多时。◀

8.13.2 MA 模型

随机序列的 MA 模型没有递归的形式；相反，它是白输入序列的有限加权和，如下所示：

$$X[k]=\sum_{n=0}^{M-1}b_nW[k-n] \tag{8-287}$$

其中 M 是阶数。这是一个 FIR 滤波器，传递函数为：

$$H(z)=\sum_{n=0}^{M-1}b_nz^{-n}\triangleq B(z) \tag{8-288}$$

系统框图如图 8-29b 所示，其功率谱密度是：

$$S_{XX}(z)=\sigma_W^2B(z)B(z^{-1}) \tag{8-289}$$

因为序列模型不是递归的，自相关函数很容易计算，一般来说：

$$\begin{aligned}\mathcal{E}[X[k]X[k+m]]&=\sum_{n=0}^{M-1}\sum_{p=0}^{M-1}b_nb_p\,\mathcal{E}[W[k-n]W[k+m-p]]\\&=\sigma_W^2\sum_{n=0}^{M-1}\sum_{p=0}^{M-1}b_nb_p\delta[m+n-p]=\sigma_W^2\sum_{n=0}^{M-m-1}b_nb_{m+n}\end{aligned} \tag{8-290}$$

其中 $m\geqslant 0$。由自相关函数的对称性，对于任意 m

$$R_{XX}[m]=\sigma_W^2\sum_{n=0}^{M-|m|-1}b_nb_{|m|+n} \tag{8-291}$$

求和实际上是滤波器参数$\{b_n\}$的确定性自相关函数。不同的 m 值使得 $b_{|m|+n}$ 相对于 b_n 进行左移或右移。因为只有 M 个参数，$R_{XX}[m]$也具有有限持续时间：当$|m|>M-1$ 时为 0，如图 8-32 所示，其中我们定义了向量 $\boldsymbol{b}\triangleq[b_0,\cdots,b_{M-1}]^{\mathrm{T}}$。图 8-33 画出了图 8-32 的自相关函数($\sigma_W^2=1$)对应的 MA 序列的现实的例子。输出序列具有显著的相关性，但相关性比前面 AR 模型的例子要低，这是由于 AR 过程有反馈：当时间滞后无穷大时，自相关函数非 0。

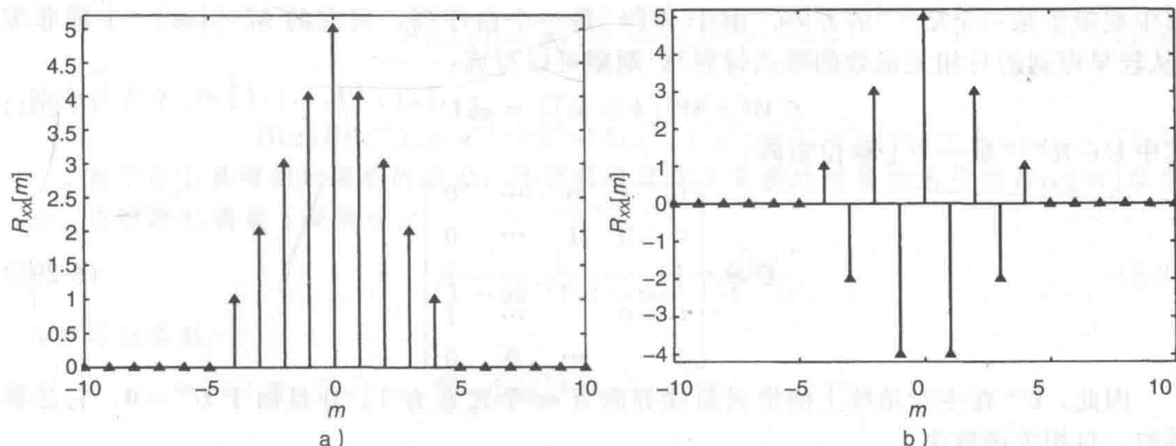


图 8-32 MA(4)序列的相关函数的例子。a) MA(4)的自相关函数，$\boldsymbol{b}=[1,1,1,1,1]^{\mathrm{T}}$；b) MA(4)的自相关函数，$\boldsymbol{b}=[1,-1,1,-1,1]^{\mathrm{T}}$

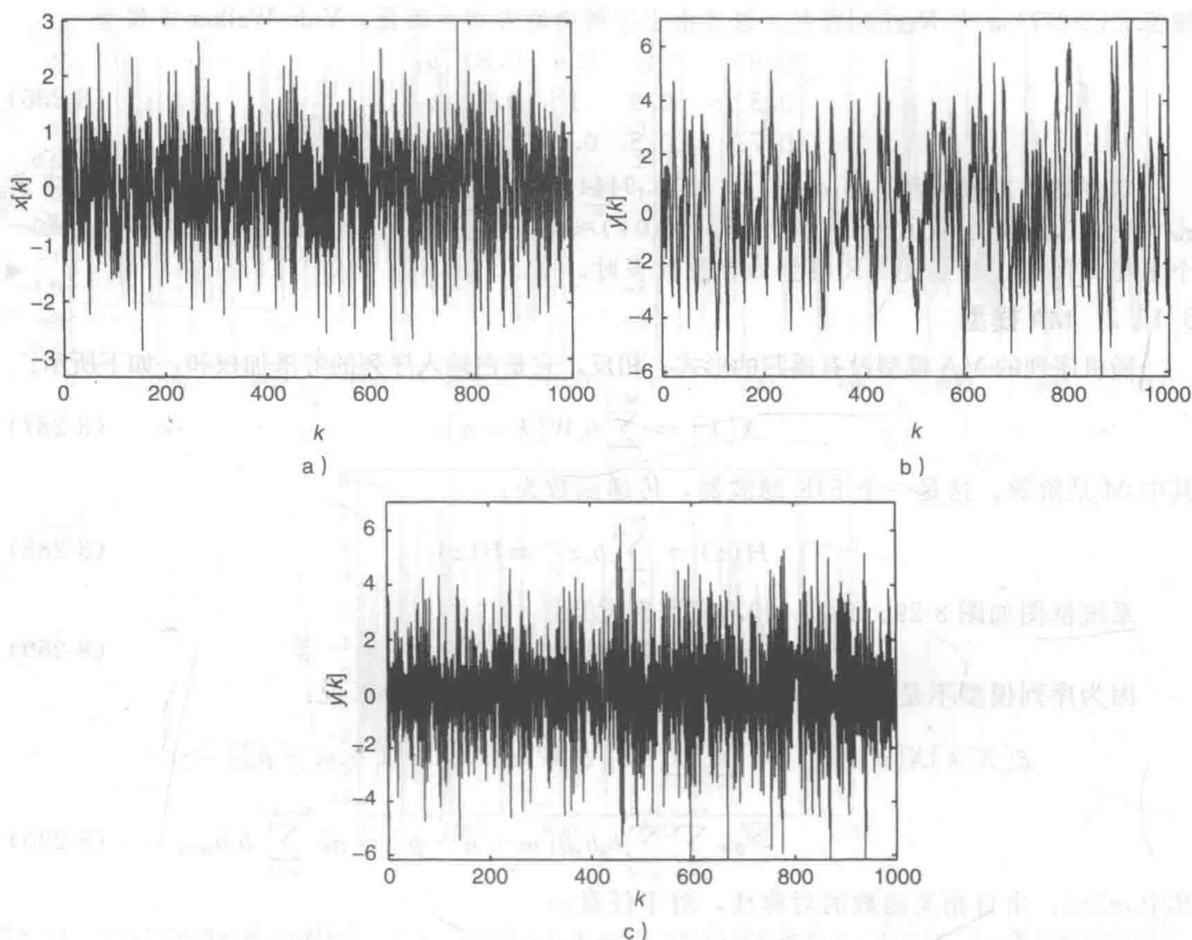


图 8-33 不同参数对应的 MA(4)序列的现实的例子。a) 白高斯输入；b) $\boldsymbol{b}=[1,1,1,1,1]^{\mathrm{T}}$ 对应的输出；c) $\boldsymbol{b}=[1,-1,1,-1,1]^{\mathrm{T}}$ 对应的输出

利用系数向量 $\boldsymbol{b}$，随机序列可以表示为以下内积：

$$X[k]=\boldsymbol{b}^{\mathrm{T}}\boldsymbol{W}[k] \tag{8-292}$$

其中 $\boldsymbol{W}[k]\triangleq[W[k],\cdots,W[k-M+1]]^{\mathrm{T}}$。这个向量表示对 FIR 滤波器很有用，在第 12 章自适应滤波中将再次出现。自相关函数可以使用这个形式写成：

$$R_{XX}[m]=\mathcal{E}[X[k]X[k+m]]=\boldsymbol{b}^{\mathrm{T}}\ \mathcal{E}[\boldsymbol{W}[k]\boldsymbol{W}^{\mathrm{T}}[k+m]]\boldsymbol{b} \tag{8-293}$$

其中期望里是一个 $\mathcal{R}^{M\times M}$ 的方阵。由于 $W[k]$ 是一个白序列，只有的 $M-|m|-1$ 项非零(从较早得到的自相关函数的形式得到)。期望可以写成：

$$\mathcal{E}[\boldsymbol{W}[k]\boldsymbol{W}^{\mathrm{T}}[k+m]]=\sigma_W^2\boldsymbol{U}^{|m|} \tag{8-294}$$

其中 $\boldsymbol{U}\in\mathcal{R}^{M\times M}$ 是一个上移位矩阵：

$$\boldsymbol{U}\triangleq\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & & \ddots & & \vdots\\ 0 & 0 & & \cdots & 1\\ 0 & & \cdots & 0 & 0\end{bmatrix} \tag{8-295}$$

因此，$\boldsymbol{U}^m$ 在主对角线上侧沿对角线方向有 m 个元素为 1，并且由于 $\boldsymbol{U}^M=\boldsymbol{0}$，它是幂零的。自相关函数为

$$R_{XX}[m]=\sigma_W^2\boldsymbol{b}^{\mathrm{T}}\boldsymbol{U}^{|m|}\boldsymbol{b} \tag{8-296}$$

这是式(8-291)的另一种表达式。收集向量里的所有自相关项，得到下列 Yule-Walker

型的方程组：

$$\begin{bmatrix} R_{XX}[0] \\ R_{XX}[1] \\ \vdots \\ R_{XX}[M-1] \end{bmatrix} = \sigma_W^2 \begin{bmatrix} \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{b} \\ \vdots \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{U}^{M-1}\boldsymbol{b} \end{bmatrix} \tag{8-297}$$

它具有有限长度，因为自相关函数在 $M-1$ 时刻以后是 0。遗憾的是，这个表达式是参数的非线性函数，一般不容易求解。

例 8-15 考虑 MA 模型的最一般情况，$M=2$，$\boldsymbol{b}=[b_0, b_1]^{\mathrm{T}}$。从而

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} = b_0^2 + b_1^2, \boldsymbol{b}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{b} = b_0 b_1 \tag{8-298}$$

令 $\sigma_W^2=1$，$R_{XX}[0]=r_0$，$R_{XX}[1]=r_1$ 将 $b_0=r_1/b_1$ 代入 $b_0^2+b_1^2=r_0$，得到

$$b_1^4 - r_0 b_1^2 + r_1^2 = 0 \tag{8-299}$$

这是 b_1^2 的二次方程。b_1^2 为实数需要 $r_0^2 \geqslant 4r_1^2$，且 $R_{XX}[m]$ 是一个有效的自相关函数。对于 $r_0^2=4r_1^2$，我们发现，$b_1=\pm\sqrt{r_0/2}$，$b_0=\pm r_1\sqrt{2/r_0}$。因此，即使在这个简单的例子中，由于式(8-297)中的二次形式，解不唯一。显然，当 M 值较大时，求解 $\boldsymbol{b}$ 的这些方程变得更难，通常会有多种解决方案。◀

8.13.3 ARMA 模型

正如名字所暗示的，ARMA 模型是 AR 和 MA 模型的组合。随机序列写成：

$$X[k] = \sum_{n=1}^{N} a_n X[k-n] + \sum_{n=0}^{M-1} b_n W[k-n] = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}[k-1] + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{W}[k] \tag{8-300}$$

其中向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 定义同上。它的传递函数是：

$$H(z) = \frac{\sum_{n=0}^{N} b_n z^{-n}}{1-\sum_{n=1}^{M-1} a_n z^{-n}} = \frac{B(z)}{1-A(z)} \tag{8-301}$$

从而功率谱密度为：

$$S_{XX}(z) = \sigma_W^2 \frac{B(z)B(z^{-1})}{[1-A(z)][1-A(z^{-1})]} \tag{8-302}$$

系统框图如图 8-29c 所示。ARMA(1，4)过程的现实例子是之前 AR(1)和 MA(4)的例子的组合，如图 8-34 所示，从中可以看出因为相关性的增加，它们有更大的结构。

例 8-16 生成图 8-34 仿真结果的模型的功率谱密度为

$$S_{XX}(z) = \sigma_W^2 \frac{B(z)B(z^{-1})}{(1-az^{-1})(1-az)} \tag{8-303}$$

其中对于 $\boldsymbol{b}^{\mathrm{T}}=[1,1,1,1]$：

$$B(z)B(z^{-1}) = z^{-3} + z^3 + 2(z^{-2}+z^2) + 3(z^{-1}+z) + 4 \tag{8-304}$$

由于分子具有相对简单的形式，可以利用双边 z 变换的时移性质得到 $R_{XX}[m]$ 每个部分。由附录 C 得到 z 变换对：

$$\frac{1}{(1-az^{-1})(1-az)} \leftrightarrow \frac{|a|^m}{1-a^2} \tag{8-305}$$

从中可以得到

$$R_{XX}[m] = \frac{\sigma_W^2}{1-a^2}[4a^{|m|} + 3(a^{|m-1|} + a^{|m+1|}) + 2(a^{|m-2|} + a^{|m+2|}) + a^{|m-3|} + a^{|m+3|}] \tag{8-306}$$

图 8-35 画出了 a 分别为正负值($\sigma_W^2=1$)时自相关函数的例子。◀

图 8-34 不同参数值对应的 ARMA(1,4) 序列的现实的例子。a) 白噪声输入；b) $\boldsymbol{a}=0.9$，$\boldsymbol{b}=[1,1,1,1,1]^{\mathrm{T}}$ 对应的输出；c) $\boldsymbol{a}=-0.9$，$\boldsymbol{b}=[1,-1,1,-1,1]^{\mathrm{T}}$ 对应的输出

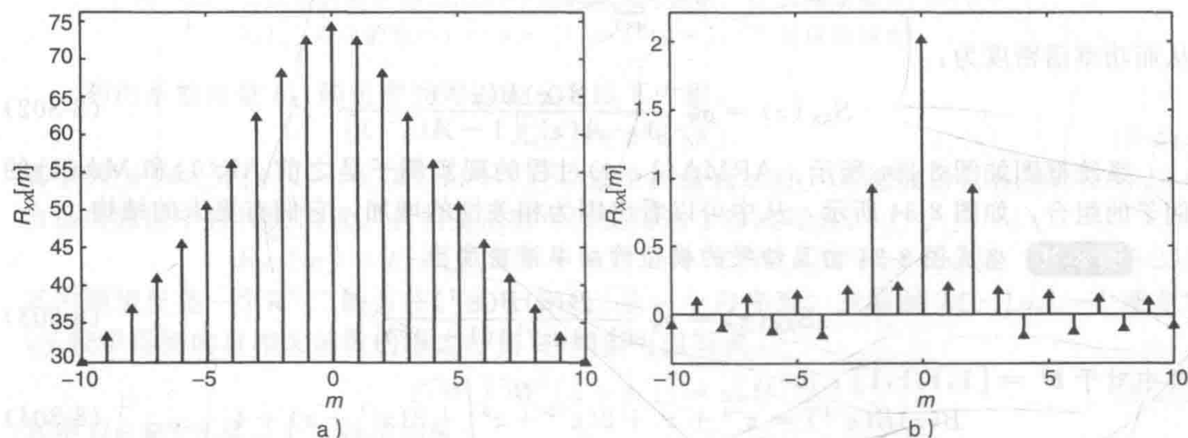


图 8-35 $\boldsymbol{b}=[1,1,1,1,1]^{\mathrm{T}}$ 时，ARMA(1,4) 的自相关函数的例子。a) $a=0.9$；b) $a=-0.9$

自相关函数可写为如下形式：

$$\mathcal{E}[X[k]X[k+m]]=\sum_{n=1}^{N}a_n\,\mathcal{E}[X[k-n]X[k+m]]+\sum_{n=0}^{M-1}b_n\,\mathcal{E}[W[k-n]X[k+m]] \tag{8-307}$$

从而得到类似滤波器本身的递归形式：

$$R_{XX}[m]=\sum_{n=1}^{N}a_nR_{XX}[m+n]+\sum_{n=0}^{M-1}b_nR_{WX}[m+n] \tag{8-308}$$

回顾输入和输出的自相关函数的关系如下：

$$R_{WX}[m] = R_{WW}[m] * h[m] \tag{8-309}$$

其中 $h[m]$ 是式(8-301)中 ARMA 传递函数的脉冲响应函数。由于 $W[k]$ 为 $R_{WW}[m]=\sigma_W^2\delta[m]$ 的白序列，互相关函数与系统的脉冲响应成正比：

$$R_{WX}[m] = \sigma_W^2 h[m] \tag{8-310}$$

使得式(8-308)变为

$$R_{XX}[m] = \sum_{n=1}^{N} a_n R_{XX}[m+n] + \sigma_W^2 \sum_{n=0}^{M-1} b_n h[m+n] \tag{8-311}$$

对于 MA 模型，等式右边第一项是 0，且第二项满足 $h[m+n]=R_{WX}[m+n]=b_{m+n}$。从而产生一个持续时间有限的自相关函数。因为(ⅰ)受等式右边第一项的影响；(ⅱ)第二项中的 $h[m]$ 实际上是 b_n 和 $h[n]$ 的确定性互相关函数，取决于 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，使得对于一般的 ARMA 模型，求解式(8-311)更为复杂。

观察发现第二个求和是不对称的：对于 $m<0$ 时，脉冲响应 $h[m+n]$ 相对于 b_n 向左移动，并且由于 $h(n)$ 为无限长，共有 M 项求和。对于 $m>0$，脉冲响应 $h[m+n]$ 相对于 b_n 向右移动，并且由于 $n\geqslant M$ 时 b_n 是零，随着右移的增加求和项的数目递减，当 $m\geqslant M$ 时，求和为 0。因此，有可能从下式计算 $\boldsymbol{a}$

$$R_{XX}[m] = \sum_{n=1}^{N} a_n R_{XX}[m+n] \tag{8-312}$$

假设 $|m|\geqslant M$，从而下面的 Yule-Walker 等式可解：

$$\begin{bmatrix} R_{XX}[M] \\ R_{XX}[M+1] \\ \vdots \\ R_{XX}[M+N-1] \end{bmatrix} = \begin{bmatrix} R_{XX}[M+1] & \cdots & R_{XX}[M+N] \\ R_{XX}[M+2] & \cdots & R_{XX}[M+N+1] \\ \vdots & & \vdots \\ R_{XX}[M+N] & \cdots & R_{XX}[M+2N] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \tag{8-313}$$

这个表达式不同于 AR 模型的式(8-282)，区别在于自相关函数的时间滞后量必须超过 ARMA 模型的 MA 部分的阶数。遗憾的是，即使能够简单得到 $\boldsymbol{a}$，仍然难以计算 $\boldsymbol{b}$，因为在式(8-311)中它乘以脉冲响应 $h[n]$，而 $h[n]$ 又取决于 $\boldsymbol{a}$ 和 $\boldsymbol{b}$；该问题的非线性度比求解 MA 模型的 $\boldsymbol{b}$ 更高。

重新整理(8-311)如下：

$$R_{XX}[m] - \sum_{n=1}^{N} a_n R_{XX}[m+n] = \sigma_W^2 \sum_{n=0}^{M-1} b_n h[m+n] \tag{8-314}$$

其中有 z 变换：

$$[1-A(z^{-1})]S_{XX}(z) = \sigma_W^2 B(z) H(z^{-1}) \tag{8-315}$$

代入式(8-301)中的 $H(z)$，并重新整理得到：

$$[1-A(z)][1-A(z^{-1})]S_{XX}(z) = \sigma_W^2 B(z) B(z^{-1}) \tag{8-316}$$

上式与式(8-302)相同。右侧的 z 反变换与 MA 模型的式(8-291)相同。因此，ARMA 模型参数可以产生如下：(ⅰ)从 $X[k]$ 现实的时间平均估计 $R_{XX}[m]$ 和 $S_{XX}[z]$；(ⅱ)使用式(8-313)的 Yule-Walker 方程计算参数 $A(z)$；(ⅲ)得到式(8-316)左边的 z 反变换；(ⅳ)求解等式右边得到 b_m。与 MA 模型情况一样，最后一步是非线性的，可以有多种解决方案。

最后，我们介绍另一种参数估计方法，假设随机序列具有特定的信号和噪声结构，使得可以得到自相关矩阵的特征结构。比起一般的谱估计，这种子空间方法更适用于噪声中的窄带信号检测。第 13 章将介绍子空间方法，如分离从不同方向的天线阵列上的得到的对应于多输入系统的多个窄带信号。

8.14 系统辨识

系统辨识与 8.13 节信号模型密切相关，如图 8-36 所示。我们感兴趣的是从它的输出 $Y[k]$ 和输入 $V[k]$ 的测量结果中估计未知系统的 $B(z)/A(z)$。我们通常假设 $V[k]$ 为白序

列，但仍假定输入可测(前面的信号建模假定它不可测)。然而，受不可测量的加性白噪声 $W[k]$的影响，系统辨识过程并不简单。此外，噪声可能被某些传递函数 $C(z)/D(z)$过滤，使得输出非白(功率谱密度不平坦)。在这里可能既要估计 $B(z)/A(z)$，又要估计 $C(z)/D(z)$。此外，图中求和得到的输出可能还要进一步通过未知滤波器 $F(z)$，造成输出 $X[k]$不能直接观察得到。假定用于滤波器组件辨识的各种算法(如果有的话)的多项式是1。最简单的情况假定 $A(z)=C(z)=D(z)=F(z)=1$，使得仅需从 $V[k]$和 $X[k]$的测量结果估计 $B(z)$。

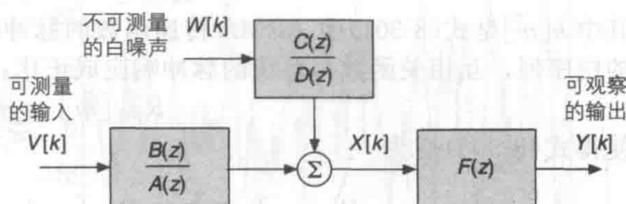


图 8-36　通用的系统辨识模型

系统辨识的目的是确定描述测量信号的滤波多项式，与信号参数建模的技术类似。信号建模中得到的系统可以反过来用于合成测量信号，常通过对白噪声序列滤波得到。该滤波器还提供了谱信息。与此相反，系统辨识的目标是产生一个模型，它还可以有其他应用，例如在控制系统中需要复杂系统模型来控制实际系统。

假设可以得到图 8-36 所示的系统模型的输入 $V[k]$和输出 $Y[k]$。注意实际系统可能不是这样的形式，有可能包含非线性和时变分量。然而，如果非线性部分影响较小，并且模型参数的计算和更新足够快，滤波器多项式可以相当准确。混合使用时域和 z 域符号，可以写成

$$F(z)Y[k]=\frac{B(z)}{A(z)}V[k]+\frac{C(z)}{D(z)}W[k] \tag{8-317}$$

其中多项式定义如下(注意其中一些第一项必须为1)：

$$A(z)\triangleq 1+a_1z^{-1}+\cdots+a_{M_a}z^{-M_a} \tag{8-318}$$

$$B(z)\triangleq b_0+b_1z^{-1}+\cdots+b_{M_b}z^{-M_b} \tag{8-319}$$

$$C(z)\triangleq c_0+c_1z^{-1}+\cdots+c_{M_c}z^{-M_c} \tag{8-320}$$

$$D(z)\triangleq 1+d_1z^{-1}+\cdots+d_{M_d}z^{-M_d} \tag{8-321}$$

$$F(z)\triangleq 1+f_1z^{-1}+\cdots+f_{M_f}z^{-M_f} \tag{8-322}$$

混合符号含义如下：

$$F(z)Y[k]=Y[k]+\sum_{m=1}^{M_f}f_mY[k-m] \tag{8-323}$$

其中这里的 z^{-1} 函数表示延时一个单位：$z^{-1}Y[k]=Y[k-1]$。

前面的模型相当普遍，并且在大多数应用中，多项式的一项或多项假定为1。我们总结了一些特殊的情况。

- 有外加输入的 ARMA 模型(ARMAX)。$A(z)=D(z)=1$：

  $$F(z)Y[k]=B(z)V[k]+C(z)W[k] \tag{8-324}$$

  ARMAX 模型的一个特例是当 $C(z)=1$ 时：

  $$F(z)Y[k]=B(z)V[k]+W[k] \tag{8-325}$$

  这被称为ARX模型。

- ARMA 模型。$A(z)=D(z)=1$，$B(z)=0$：

  $$F(z)Y[k]=C(z)W[k]\Rightarrow Y[k]=\frac{C(z)}{F(z)}W[k] \tag{8-326}$$

  特殊情况包括 MA 模型($F(z)=1$)：

  $$Y(k)=C(z)W[k] \tag{8-327}$$

  和 AR 模型($C(z)=1$)：

  $$F(z)Y(k)=W[k]\Rightarrow Y[k]=\frac{1}{F(z)}W[k] \tag{8-328}$$

● Box-Jenkins 模型。$F(z)=1$：

$$Y[k]=\frac{B(z)}{A(z)}V[k]+\frac{C(z)}{D(z)}W[k] \tag{8-329}$$

● 输出误差。$F(z)=1$ 且 $C(z)=D(z)$：

$$Y[k]=\frac{B(z)}{A(z)}V[k]+W[k] \tag{8-330}$$

这对应于有加性噪声的 IIR 滤波器，当 $A(z)=1$ 时变成一个 FIR 滤波器。FIR 和 IIR 输出误差滤波器自适应算法将在第 12 章讨论。

● 公式误差。$A(z)=1$ 且 $C(z)=D(z)=1$：

$$F(z)Y[k]=B(z)V[k]+W[k] \tag{8-331}$$

公式误差滤波器自适应算法也将在第 12 章讨论。

图 8-37 中给出了输出误差和公式误差模型，加性噪声没必要是白序列；因此，我们上面可以包括 $C(z)\neq D(z)$，但第 12 章没有像 Box-Jenkins 模型那样辨识噪声多项式。

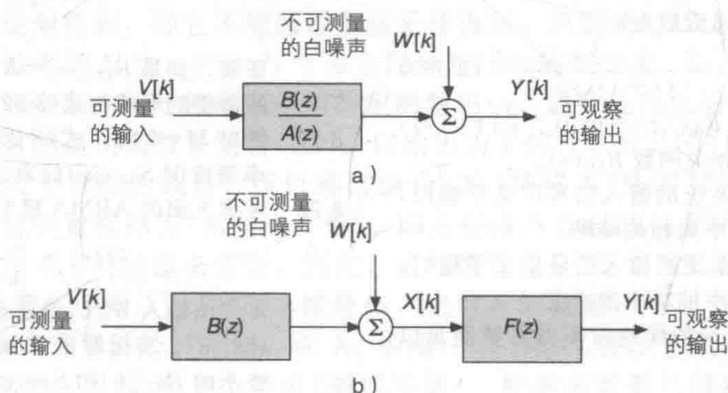


图 8-37 IIR 系统辨识。a）输出误差模型；b）公式误差模型

习题

各态历经性

8-1 (a)推导式(8-23)的随机序列 $X[k]$具有均值各态历经性的必要和充分条件；(b)证明式(8-24)的充分条件。

8-2 设 $X[k]$均值为 0，自相关函数 $C_{XX}[m]=a^{|m|}$。判断该随机序列是否满足习题 8-1 的两个条件。

8-3 假设 $X(t)$具有均值各态历经性，Y 是一个与 $X(t)$独立的随机变量。定义随机过程 $Z(t)=YX(t)$，判断该过程是否也具有均值各态历经性。注意一个现实的形式是 $z(t)=yx(t)$，因此当 $x(t)$变化时，y 取值不变。

8-4 随机过程 $X(t)$均值为 0，自相关函数如下：

$$R_{XX}(\tau)=(1-\tau^2)I_{[-1,1]}(\tau) \tag{8-332}$$

判断 $X(t)$是否具有均值各态历经性。

8-5 随机过程 $X(t)$具有以下周期自相关函数：

$$C_{XX}(\tau)=2\cos(2\pi f_c\tau) \tag{8-333}$$

判断当频率 f_c 固定时，$X(t)$是否具有均值各态历经性。

8-6 如果 $R_{XX}(\tau)=\exp(-|\tau|)$，利用例 8-3 的结果判断该零均值高斯随机过程 $X(t)$是否具有相关函数各态历经性。

功率谱密度

8-7 对于互功率谱密度，证明：(a) $S_{XY}(f)=S_{YX}(-f)$；(b)虚部是奇函数。

8-8 考虑下列随机过程：

$$X(t)=a\sin(2\pi f_c t+\Phi) \tag{8-334}$$

其中 Φ 是$[-\pi/2,\ \pi/2]$上的均匀分布。计算：(a)自相关函数 $R_{XX}(\tau)$；(b)功率谱密度$S_{XX}(f)$。

8-9 (a)计算下式的功率谱密度和频率范围 $f\in[f_c-W,f_c+W]$上的平均功率。其中 $W\ll f_c$

$$R_{XX}(\tau)=\exp(-\lambda|\tau|)\cos(2\pi f_c\tau) \tag{8-335}$$

(b)写出功率谱分布的表达式。

8-10 对

$$R_{XX}(\tau)=\exp(-\tau^2/2\sigma^2) \tag{8-336}$$

重复习题 8-9。其中 $f\in[-W,W]$，$W=\sigma$。

8-11 对习题 8-9 中的 $R_{XX}(\tau)$进行功率谱密度分解，$S_{XX}(s)=S_{XX}^{+}(s)S_{XX}^{-}(s)$。

8-12 设 $X(t)=S(t)+W(t)$，其中 $S(t)$和 $W(t)$的自相关函数分别为 $R_{SS}(\tau)=5\exp(-|\tau|)$和 $R_{WW}(\tau)=\delta(\tau)$。假设 $S(t)$和 $W(t)$不相关，

对$X(t)$的功率谱密度进行因子分解。

8-13 设$X[k]$是一个广义平稳随机序列，均值为0，自相关函数$R_{XX}[m]=(N_0/2)\delta[m]$。证明为得到平坦功率谱密度，对应的白噪声序列均值必须为零。

线性时不变系统

8-14 线性时不变系统满足以下差分方程：

$$Y[k]=X[k]+(1/2)X[k-1]+(1/4)Y[k-1] \quad (8\text{-}337)$$

如果输入的$R_{XX}[m]=2\delta[m]$，计算输出的功率谱密度$S_{YY}(j\omega)$。

8-15 计算下列系统的功率谱密度：

$$Y(t)=\int_0^t X(u)du \quad (8\text{-}338)$$

其中$R_{XX}(\tau)=2\delta(\tau)$

8-16 线性系统的传递函数为：

$$H(s)=\frac{2}{(s+1)(s+2)} \quad (8\text{-}339)$$

当输入信号的$R_{XX}(\tau)=\exp(-2|\tau|)$时，计算输出的自相关函数$R_{YY}(\tau)$。

8-17 当线性时不变系统的输入信号广义平稳时，写出式(8-152)中的相关结果。

8-18 当线性时不变系统的输入信号广义平稳时，写出式(8-156)中的功率谱密度。

8-19 输入信号为$X(t)$的线性时不变系统满足以下微分方程：

$$\frac{d^2Y(t)}{dt^2}+4\frac{dY(t)}{dt}+2Y(t)=X(t) \quad (8\text{-}340)$$

其中$R_{XX}(\tau)=(1/2)\delta(\tau)$。写出下列表达式：(a)$S_{XY}(f)$和$R_{XY}(\tau)$；(b)$S_{YY}(f)$和$R_{YY}(\tau)$。

8-20 考虑一个线性时不变系统，其输入为$X(t)$，输出为$Y(t)$，冲激响应

$$h(t)=t\exp(-\lambda t)u(t) \quad (8\text{-}341)$$

其中$\lambda>0$。当$R_{XX}(\tau)=\delta(\tau)$时，写出$R_{YX}(\tau)$和$R_{YY}(\tau)$的表达式。

8-21 假设$X_1[k]$经$H_1(z)$滤波后的输出为$Y_1[k]$，$X_2[k]$经$H_2(z)$滤波后的输出为$Y_2[k]$。(a)写出$S_{Y_1Y_2}(z)$的$S_{X_1X_2}(z)$形式的表达式；(b)如果系统的频率响应互不重叠(例如带通系统)，证明输出过程是正交的。

带宽

8-22 计算下列功率谱密度的：(a)3dB带宽；(b)噪声等效带宽：

$$S_{XX}(f)=4(1-|f|/5)I_{[-5,5]}(f) \quad (8\text{-}342)$$

8-23 如果

$$S_{XX}(f)=\frac{\sin^2(2\pi f)}{\pi^2 f^2} \quad (8\text{-}343)$$

重复习题8-22

8-24 计算习题8-10对应的功率谱密度的：(a)均方根带宽；(b)噪声等效带宽。

谱估计

8-25 利用离散时间傅里叶变换的性质，证明式(8-225)和式(8-227)都是周期图的等价表达式。

8-26 考虑互相关函数$R_{XY}[m]$的下列估计：

$$\hat{R}_{XY}[m]=\frac{1}{N}\sum_{k=0}^{N-|m|-1}x[k]y[k+m] \quad (8\text{-}344)$$

证明，如果$R_{YX}[m]=R_{XY}[-m]$，这个估计的期望的形式与式(8-223)类似。

8-27 导出$M=2$时，式(8-287)中MA序列的功率谱密度$S_{XX}(j\omega)$的表达式。

8-28 考虑下面的ARMA模型：

$$H(z)=\frac{b_0+b_1z^{-1}}{1-a_1z^{-1}} \quad (8\text{-}345)$$

零均值输入$W[k]$的自相关函数$R_{WW}[m]=\sigma_W^2\delta[m]$。写出输出$X[k]$用参数$\{a_1,b_0,b_1\}$表示的$R_{XX}[k]$和$S_{XX}(z)$。

8-29 对于具有N个参数的一般AR模型，当输入$W[k]$是方差为σ_W^2的白噪声序列时，推导输出$X[k]$用$R_{XX}[m]$表示的方差。

仿真作业

8-30 用randn产生1000个方差为1的高斯白噪声$W[k]$的样本，并使用下列滤波器得到输出$X[k]$：

$$H(z)=\frac{2}{1-(5/6)z^{-1}+(1/6)z^{-2}} \quad (8\text{-}346)$$

画出$W[k]$和$X[k]$的现实例子。在计算过程$X[k]$的周期图时，先采用矩形窗，再利用Bartlett窗口，画出得到的周期图。

8-31 对于习题8-30的信号，利用滞后{0，1，2，3，4}的xcorr，推导自相关函数$R_{XX}[m]$的估计。利用MATLAB，将Yule-Walker方程中的N取2～4，估计这个系统的参数。

进一步阅读

本章涉及的资料也可以从第6章末尾和第7章随机过程对应的参考文献中找到。有关特定主题的详细信息可在以下文献找到。噪声：Cohen(2005)、Larson和Shubert(1979)。谱估计：Kay和Marple(1981)，Kay(1986)，Porat(1994)，Stoica和Moses(1997)。信号参数模型和系统辨识：Box和Jenkins(1970)，Chatfield(1980)，Ljung和Söderström(1983)，Ljung(1987)，Therrien(1992)。

第9章 充分统计量和参数估计

9.1 引言

本章，我们想找到由一组随机变量组成的函数，以估计基本分布中的未知参数。这样的函数被称为统计量，它通常会在不丢失参数“信息”的前提下将随机变量的数目(被称为样本)减少到最小。这个过程可总结如下：

- 设 $X(t)$是一个随机过程，具有参数为标量 θ 的固定的概率密度函数 $f_X(x)$。假设 θ 无法通过观测得到，即它不能通过直接采样得到，只能通过 $X(t)$的样本得到。
- 随机过程被采样 N 次，产生 N 个独立同分布(iid)随机变量 $\{X_1,\cdots,X_N\}$。尽管$\{X_n\}$包含了未知参数 θ 同等的信息，但特定的输出$\{x_n\}$通常是变化的。
- 采样后的目的是用函数来组合$\{X_n\}$，以给出为了估计 θ 所需的“更简单”的量。由于 N 通常远大于参数的数目，所以通过组合$\{X_n\}$将样本“浓缩”到一个或多个随机变量，该随机变量被称为“充分统计量”。因为在将 N 个随机变量映射到较少随机变量的过程中关于 θ 的信息未丢失，因此，这些统计量是“充分”的。

$\{X_n\}$也可从随机序列 $X[k]$的样本推导出，它也可能就是一个不考虑时间的特定随机变量的 N 个样本。为了符号的方便，在 X_n 中用一个下标来表示它是随机过程 $X(t)$、随机序列 $X[k]$或随机变量 X 的一个样本(随机变量)。样本不需要从过程/序列的均匀时间抽样中收集；我们可以简单的将$\{X_n\}$看做一组参数为 θ 的 iid 随机变量。

为了简化方程，用 $\boldsymbol{X}\triangleq\{X_1,\dots,X_N\}$ 来代表 N 个样本，$\boldsymbol{x}\triangleq\{x_1,\cdots,x_N\}$ 来代表特定输出。(注意到这里用黑斜体表示的记号也用在标示随机向量和它们的输出。)

9.2 统计量

定义(统计量) **统计量** T 是随机变量 $\boldsymbol{X}$ 的函数，它本身也是一个随机变量。

关于符号，$\boldsymbol{X}$ 的函数是非随机的，我们可以写为 $T=g(\boldsymbol{X})$，但这也不是必须的，本章后面部分才需要。对于特定的输出 $\boldsymbol{X}=\boldsymbol{x}$，统计量的值是 $T=t$。样本 T 通常被用在估计器的统计描述中(所以我们在这章一开始就用它)，注意不要与用来表示时间间隔的 T 混淆。

定义(估计器) θ 的一个**估计器**是一个统计量 T，它是样本 $\boldsymbol{X}$ 的函数，并且具有特定的理想特性。

例 9-1 样本均值可能是运用最广泛的统计量：

$$T_1=\frac{1}{N}\sum_{n=1}^{N}X_n\triangleq\overline{X} \tag{9-1}$$

它是实际均值 μ_X 的无偏估计器。以下统计量

$$T_2=\sum_{n=1}^{N}X_n \tag{9-2}$$

包括对于 μ_X 与 $\overline{X}$ 基本相同的信息，但它是一个有偏估计器。T_1 和 T_2 在统计意义上是等价的，因为它们可由一一对应的函数 $g(\cdot)$联系起来：

$$T_1=g(T_2)=\frac{1}{N}T_2,\quad T_2=g^{-1}(T_1)=NT_1 \tag{9-3}$$

其中 $g^{-1}(\cdot)$是反向映射。稍后，我们将深入讨论统计量是样本的函数或映射这个观念。◀

例 9-2 样本方差是一个统计量：

$$T=\frac{1}{N-1}\sum_{n=1}^{N}(X_n-\overline{X})^2\triangleq S^2 \tag{9-4}$$

求和项除以 $N-1$(而不是 N，这是"自然的"选择)，这样才能保证样本方差是真实方差的无偏估计器。◀

9.3 充分统计量

充分统计量对于参数估计问题非常有用。考虑具有未知参数 θ 的 pdf $f_X(x;\theta)$，符号强调了 θ 是待估参量。例如，θ 可以是参数为 $\{\mu,\sigma\}$ 的 iid 高斯随机序列的均值和方差。假设我们有 N 个样本 $\boldsymbol{X}$，T 为样本的函数(映射)。初始的 N 个观测值构造了充分统计量 T，它包括了就估计 θ 而言相同的信息：对于实现估计它是充分的。因此，如果 T 是可用的并且充分的，那么初始的观测数据集 $\boldsymbol{X}$ 并不会对估计 θ 提供更多的信息。

定义(充分统计量) 统计量 T 在估计参数 θ 时是充分的，如果条件概率 $P(\boldsymbol{X}=\boldsymbol{x}|T=t)$ 不是 θ 的函数。当随机变量是连续的，如果条件 pdf $f_{\boldsymbol{X}|T}(\boldsymbol{x}|t)$ 不是 θ 的函数，则 T 是充分的。

条件概率可被重写为：

$$P(\boldsymbol{X}=\boldsymbol{x}|T=t)=\frac{P(\boldsymbol{X}=\boldsymbol{x},T=t)}{P(T=t)}=\frac{P(\boldsymbol{X}=\boldsymbol{x})}{P(T=t)} \tag{9-5}$$

第二个等式成立是因为 $\{\boldsymbol{x},t\}$ 与 $\boldsymbol{x}$ 等价：我们可由 $\boldsymbol{x}$ 确定 t。最后一个表达式中的分子和分母都与 θ 相关，尽管式中并未明确显示。分子是输出 $\boldsymbol{x}$ 的联合概率，分母是具有输出 t 的统计量的概率(也是个随机变量)。充分统计量的定义的隐含意义是这两个概率同样依赖于 θ，因此分子和分母中 θ 的函数可约掉。图 9-1 中的方框图总结了参数估计中用到的充分统计量。

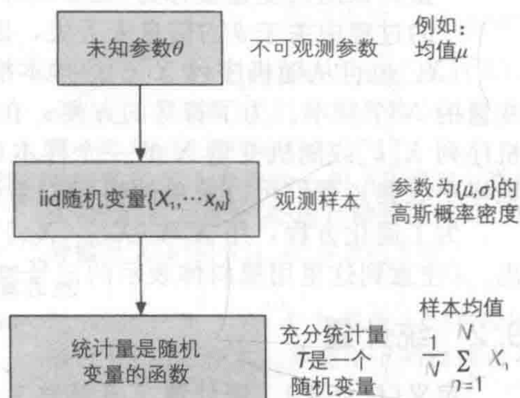


图 9-1 参数估计中用到的充分统计量的总结

一个充分统计量可以是多维的，即使对于标量参数 θ。后面，我们将定义一个最小充分统计量，它由统计量的维数定义，还有一个完备充分统计量，它的定义更复杂些。

例 9-3 设 $\{X_n\}$ 是一个参数为 p 的伯努利随机变量的样本，求和量 $T=\sum\limits_{n=1}^{N}X_n$ 具有二项式分布，参数为 $\{N,p\}$，由式(9-5)：

$$\frac{P(\boldsymbol{X}=\boldsymbol{x})}{P(T=t)}=\frac{\prod\limits_{n=1}^{N}p^{x_n}q^{1-x_n}}{\binom{N}{t}p^tq^{N-t}}=\frac{p^{\sum_{n=1}^{N}x_n}q^{N-\sum_{n=1}^{N}x_n}}{\binom{N}{t}p^tq^{N-t}}=\frac{1}{\binom{N}{t}} \tag{9-6}$$

得到最后的结果是因为 $t=\sum\limits_{n=1}^{N}x_n$，由于二项式系数不是 p 的函数(或 $q\triangleq 1-p$)，T 对 p 来说是充分的。◀

例 9-4 设 $\{X_n\}$ 是一个参数 $\theta=\lambda$ 的指数分布，因此它们的联合 pdf 为：

$$f_{\boldsymbol{X}}(\boldsymbol{x})=\lambda^N\exp\left(-\lambda\sum_{n=1}^{N}x_n\right)\prod_{n=1}^{N}I_{[0,\infty)}(x_n) \tag{9-7}$$

为了确定 $T=\sum\limits_{n=1}^{N}X_n$ 是否充分，考虑以下随机变量集合的变换，这样计算需要的条件

pdf 更简单：对 $n=1,\cdots,N-1, Y_n=X_n, Y_N=\sum_{n=1}^{N}X_n$。这个变换是可逆的，因为 $X_N=Y_N-\sum_{n=1}^{N-1}Y_n$。$\boldsymbol{Y}\triangleq[Y_1,\cdots,Y_N]^T$ 的联合 pdf 为：

$$f_{\boldsymbol{Y}}(\boldsymbol{y})=\lambda^N\exp(-\lambda y_N)I_{[0,y_N]}\left(\sum_{n=1}^{N-1}y_n\right)\prod_{n=1}^{N}I_{[0,\infty)}(y_n) \tag{9-8}$$

其中的指示函数确保了对于非负的 $\{y_n\}$ 和 $y_N\geqslant\sum_{n=1}^{N-1}y_n$，pdf 是非零的。之所以选择这种特定的变换是因为 Y_N 是待考察的统计量。我们强调 $\{Y_n\}$ 的联合 pdf 与 $\{X_n\}$ 的不同。但由于变换是可逆的，可证明 Y_N 是对 λ 来说是充分的，也就证明 $\sum_{n=1}^{N}X_n$ 是充分的。条件 pdf 为：

$$f_{Y_1,\cdots,Y_{N-1}|Y_N}(y_1,\cdots,y_{N-1}|y_N)=\frac{f_{\boldsymbol{Y}}(\boldsymbol{y})}{f_{Y_N}(y_N)}=\frac{\lambda^N\exp(-\lambda y_N)}{(\lambda^N/\Gamma(N)]y_N^{N-1}\exp(-\lambda y_N)}$$

$$\cdot I_{(0,y_N]}\left(\sum_{n=1}^{N-1}y_n\right)\prod_{n=1}^{N}I_{[0,\infty)}(y_n) \tag{9-9}$$

其中分母 $f_{Y_N}(y_N)$ 是参数为 $\{\lambda,\ r=N\}$ 的一个伽马随机变量的 pdf。化简后，我们看到条件 pdf 与 λ 独立：

$$f_{Y_1,\cdots,Y_{N-1}|Y_N}(y_1,\cdots,y_{N-1}|y_N)=[\Gamma(N)/y_N^{N-1}]I_{[0,y_N]}\left(\sum_{n=1}^{N-1}y_n\right)\prod_{n=1}^{N}I_{[0,\infty)}(y_n) \tag{9-10}$$

因此 $T=Y_N$ 对于 λ 是充分的。◀

尽管一个充分统计量的定义很具有吸引力，但并不总能很容易的推导出条件概率：用定义来证明一个统计量是充分的是非常冗长乏味的。接下来的定理允许只用检查样本的联合 pdf 来确定统计量是否充分，不再需要将其重写为一个条件 pdf。

定理 9-1（费希尔-纽曼因子分解）　设 $\boldsymbol{X}$ 是从具有未知参量 θ 的 pdf $f_X(x;\ \theta)$ 中得到的样本。统计量 T 对于 θ 是充分的，当且仅当它们的联合 pdf 可做如下分解：

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta)=h(\boldsymbol{x})g(t;\theta) \tag{9-11}$$

其中 $h(\boldsymbol{x})$ 是一个非负函数，与 θ 无关，$g(t;\ \theta)$ 是一个与 θ 相关的非负函数，且只通过 $T=t$ 才是输出 $\boldsymbol{x}$ 的函数。

证明：必要性：假设 T 是一个充分统计量，我们可以得到：

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta)=f_{\boldsymbol{X},T}(\boldsymbol{x},t;\theta) \tag{9-12}$$

等式成立是因为 T 是一个观测的函数，并且除了初始样本的信息外并没有提供其他信息。证明可从联合累积分布函数（cdf）开始，然后证明因为包括了 T，概率不会改变，与式(9-5)所做类似。将此表达式重写为条件概率的形式：

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta)=f_{\boldsymbol{X}|T}(\boldsymbol{x}|t;\theta)f_T(t;\theta)=h(\boldsymbol{x})g(t;\theta) \tag{9-13}$$

其中 $h(\boldsymbol{x})\triangleq f_{\boldsymbol{X}|T}(\boldsymbol{x}|t;\ \theta)$ 与 θ 无关，因为它以 T 为条件，而我们已假设 T 是充分的。另一项 $g(t;\ \theta)\triangleq f_T(t;\ \theta)$ 是一个 θ 的函数，只通过 t 与样本相关。

充分性：如果统计量的 pdf 为：

$$f_T(t;\theta)=\sum_{\boldsymbol{x}:T=t}\cdots\sum f(\boldsymbol{x},t;\theta)=\sum_{\boldsymbol{x}:T=t}\cdots\sum f(\boldsymbol{x};\theta) \tag{9-14}$$

写成了联合 pdf 的 N 个变量的和的形式。第二个表达式成立是因为 T 是一个 $\boldsymbol{X}$ 的函数。代入式(9-11)，将 $g(t;\theta)$ 提取出来得到：

$$f_T(t;\theta)=\sum_{\boldsymbol{x}:T=t}\cdots\sum h(\boldsymbol{x})g(t;\theta)=g(t;\theta)\sum_{\boldsymbol{x}:T=t}\cdots\sum h(\boldsymbol{x}) \tag{9-15}$$

由充分统计量的定义，我们需要证明以下条件 pdf 独立于 θ：

$$f_{\boldsymbol{X}|T}(\boldsymbol{x}|t;\theta)=\frac{f_{\boldsymbol{X},T}(\boldsymbol{x},t;\theta)}{f_T(t;\theta)}=\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_T(t;\theta)}$$

$$=\frac{h(\boldsymbol{x})g(t;\theta)}{g(t;\theta)\sum\limits_{\boldsymbol{x}:T=t}\cdots\sum h(\boldsymbol{x})}=\frac{h(\boldsymbol{x})}{\sum\limits_{\boldsymbol{x}:T=t}\cdots\sum h(\boldsymbol{x})} \tag{9-16}$$

第一个表达式是一个条件 pdf 的定义，第二个表达式由式(9-12)得到，第三个表达式将式(9-11)代入得到。由于包含 θ 的项被消掉，最终的结果可证明条件 pdf 不是一个 θ 的函数，证明完毕。

例 9-5 由式(9-7)中指数分布的联合 pdf，我们得到

$$h(\boldsymbol{x})=\prod_{n=1}^{N}I_{[0,\infty)}(x_n),\quad g(t;\lambda)=\lambda^N\exp(-\lambda t) \tag{9-17}$$

因此 $T=\sum\limits_{n=1}^{N}X_n$ 对于 λ 是充分的。观察到指示函数包括在 $h(\boldsymbol{x})$ 中，因为它们与 λ 无关。◀

例 9-6 当定义域有限时，指示函数尤其有用。假设各 pdf 在$[0,\theta]$上均匀分布。则联合 pdf 为：

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta)=\frac{1}{\theta^N}\prod_{n=1}^{N}I_{[0,\theta]}(x_n)-\frac{1}{\theta^N}I_{[0,\theta]}(x_{(n)}) \tag{9-18}$$

它只与 N 阶统计量(见第 4 章)相关。如果 $x_{(N)}\leqslant\theta$，则所有其他样本都小于 θ，因此其他的 $N-1$ 个指示函数可以忽略。则

$$h(\boldsymbol{x})=1,\quad g(t;\theta)=\frac{1}{\theta^N}I_{[0,\theta]}(t) \tag{9-19}$$

对于 λ，$T=X_{(N)}$ 是充分的。◀

在讨论例 9-7 前，我们来对充分统计量和顺序统计量做一个对比。由充分统计量的定义可知完整的样本 $\{X_1,\cdots,X_N\}$ 总是充分的。当然，这样选统计量没有减少变量的数目。同样，顺序统计量 $\{X_{(1)},\cdots,X_{(N)}\}$ 也是充分的。这很容易看到，因为 N 个边缘 pdf 乘积的排序不会改变联合 pdf：

$$\prod_{n=1}^{N}f_{X_n}(x_n)=\prod_{n=1}^{N}f_{X_{(n)}}(x_{(n)}) \tag{9-20}$$

因此可以得到同样的因式分解结果。由定义还可以证明顺序统计量是充分的。$\{x_n\}$的所有排序集都具有基数 $N!$ 因此

$$P(\boldsymbol{X}=\boldsymbol{x}\,|\,X_{(1)}=x_{(1)},\cdots,X_{(N)}=x_{(N)})=1/N! \tag{9-21}$$

假设 $\boldsymbol{x}$ 是一种可能的排序。由于这个条件 pdf 与参数 θ 独立(这里不需要指定)，顺序统计量是充分的。

例 9-7 设$\{X_n\}$是参数为 $\{\theta=\mu,\sigma=1\}$ 高斯序列

$$f_{\boldsymbol{X}}(\boldsymbol{x};\mu)=\frac{1}{(2\pi)^N}\exp\Big(-(1/2)\sum_{n=1}^{N}(x_n^2-2\mu x_n+\mu^2)\Big) \tag{9-22}$$

这里为了方便，我们没有用指示函数，因为每个 x_n 的定义域都是$\mathcal{R}$。观察到

$$f_{\boldsymbol{X}}(\boldsymbol{x};\mu)=\frac{1}{(2\pi)^N}\exp\Big(-(1/2)\sum_{n=1}^{N}x_n^2\Big)\exp\Big(\mu\sum_{n=1}^{N}x_n\Big)\exp(-N\mu^2/2) \tag{9-23}$$

因此，因式分解得到：

$$h(\boldsymbol{x})=\frac{1}{(2\pi)^N}\exp\Big(-(1/2)\sum_{n=1}^{N}x_n^2\Big),\,g(t;\mu)=\exp(\mu t)\exp(-N\mu^2/2) \tag{9-24}$$

因此对于 μ，$T=\sum\limits_{n=1}^{N}X_n$ 是充分的。但注意到(对于 N 为偶数)

$$\boldsymbol{T}=\Big[\sum_{n=1}^{N/2}X_n,\ \sum_{n=N/2+1}^{N}X_n\Big]^{\mathrm{T}} \tag{9-25}$$

对于 μ，它也是充分的，其中 $\boldsymbol{T}$ 为一个向量。还可以通过将这两个和式分成4个和式等方法来得到 μ 的其他充分统计量。尽管式(9-25)和类似的多维表达式对 μ 是充分的，它们却并不是我们稍后将讨论的最小充分统计量。◀

上一个例子要说明的是对于未知参数 θ 通常有多个充分统计量。表9-1总结了之前具有未知方差的高斯情况的几个充分统计量；也包括了几个统计量的例子，它们对于 μ 不是充分的。寻找充分统计量的目地就是将维数减到尽可能的小，但仍保留了充分性(关于 θ 的信息没有丢失)。最低维数的充分统计量被称为最小充分统计量。对例9-7中的高斯分布，$T=\sum_{n=1}^{N}X_n$ 显然是最小充分统计量，因为 N 个样本已减少到了一个单独的统计量。对于一个标量参数 θ，如果能够找到一个单一的充分统计量，则其自动成为最小充分统计量。因式分解定理通常能提供一个最小充分统计量，但并不总是象9-4节讨论的那样。我们用一个不可能找到一维充分统计量的标量 θ 的例子来结束这节的内容。

表9-1 参数为 $\{\theta=\mu,\sigma=1\}$ 的高斯 $\{X_n\}$ 的统计量

统计量 T	类型	维数
样本：$\boldsymbol{X}$	充分的	N
顺序：$\{X_{(1)},\cdots,X_{(N)}\}$	充分的	N
$\left[\sum_{n=1}^{N/2}X_n,\ \sum_{n=N/2+1}^{N}X_n\right]^{\mathrm{T}}$	充分的	2
求和：$\sum_{n=1}^{N}X_n$	最小充分的	1
$\{X_1,\cdots,X_{N-1}\}$(缺少一个样本)	不是充分的	$N-1$
平方求和：$\sum_{n=1}^{N}X_n^2$(缺少符号信息)	不是充分的	1

例9-8 考虑一个均匀分布的情况，具有如下的联合pdf：

$$f_{\boldsymbol{X}}(\boldsymbol{x})=\prod_{n=1}^{N}I_{[\theta-1/2,\theta+1/2]}(x_n)=I_{[\theta-1/2,\theta+1/2]}(x_{(1)})I_{[\theta-1/2,\theta+1/2)}(x_{(N)}) \tag{9-26}$$

从中得到 $\boldsymbol{T}=[X_{(1)},\ X_{(N)}]^{\mathrm{T}}$ 是充分的。直观上，显然其他的顺序统计量不需要了，因为我们只对pdf的边界感兴趣，如图9-2所描述的，对此只有 $X_{(1)}$ 和 $X_{(N)}$ 是必须的。因此，即使只有一个单一的参数 θ，最小充分统计量的维度也是2；不可能找到一个充分的单一统计量。对于 θ，$X_{(1)}$ 或 $X_{(N)}$ 各自都不是充分的，它们的和也不是。

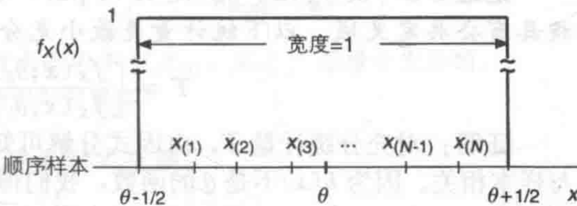


图9-2 对例9-8中均匀pdf $f_X(x)=I_{[\theta-1/2,\theta+1/2]}(x)$ 的输出示例，其中 $\{X_{(1)},\ X_{(N)}\}$ 是 θ 的最小充分统计量。(注意到纵轴中“断”线是为了使图看起来紧凑些。)◀

9.4 最小充分统计量

为了估计 θ，我们可用任意的充分统计量，但用最低维的充分统计量更实用。由于所有的统计量都是样本的函数，最小充分统计量定义如下。

定义(最小充分统计量) *参数 θ 的充分统计量 T 是最小的，如果它可以被表示为 θ 的每个其他充分统计量的函数。*

由于 N 个样本 $\{X_1,\cdots,X_N\}$ 的函数不能增加维数(定义一个函数时对每个自变量的值只有唯一的值与它对应)，显然最小充分统计量必须是所有充分统计量中最低维的。如果最小充分统计量与另一个充分统计量通过一个一一对应函数相关，则它们都是最小的，即，它们等价。注意到以上的定义中有两点需求非常重要：(ⅰ)T 必须是充分的；(ⅱ)T

是θ的所有其他充分统计量的函数。

例 9-9 对于均匀 pdf $f_X(x)=I_{[\theta-1/2,\theta+1/2]}(x)$，以下的单一统计量不是最小充分的：

$$T=(1/2)[X_{(1)}+X_{(N)}] \tag{9-27}$$

即使T是充分统计量$\{X_{(1)}, X_{(N)}\}$的一个函数，它也不是充分的，因为θ的一些信息被丢失了。由图 9-2 直觉上可将最小值和最大值平均后估计θ，它是 pdf 的中心(θ是均值)。但是，由因式分解定理，我们看到不可能将联合 pdf 中剩余的两个指示函数结合以实现$X_{(1)}$和$X_{(N)}$的求和。当$N\to\infty$，可以证明式(9-27)中的统计量是一个对θ的合理估计器。事实上，它是无偏的：

$$(1/2)\,\mathcal{E}[X_{(1)}+X_{(N)}]=\theta \tag{9-28}$$

因为对这种均匀 pdf 可以直接得到(见习题 9-4)：

$$\mathcal{E}[X_{(1)}]=\theta-1/2+\frac{1}{N+1} \tag{9-29}$$

$$\mathcal{E}[X_{(N)}]=\theta+1/2-\frac{1}{N+1} \tag{9-30}$$

当$N\to\infty$，一阶统计量的均值接近均匀 pdf 的左边界，而N阶统计量的均值接近右边界。期望值绝不会超过边界值，因为$\mathcal{E}[X_{(1)}]$中加上了$1/(N+1)$，而$\mathcal{E}[X_{(N)}]$减去了$1/(N+1)$。尽管式(9-27)看起来是个θ的不错的估计器，而且它是最小统计量(是一维的)，但它不是充分的。◀

初始的样本$\{X_1,\cdots,X_N\}$永远不可能是最小充分的。但由于顺序统计量总是充分的，而它们是初始样本的函数，可以是最小的，这就排除了初始样本是最小的。尽管不是很明显，顺序统计量实际上是初始样本的简化(即便维度仍为N)，因为初始样本的索引已经不再已知。初始样本不可能表示为顺序统计量的函数。

接下来，我们讨论两种得到最小充分统计量的方法：一种用于具有公共定义域 pdf，另一种是没有公共定义域的。一个含未知均值μ的高斯随机变量有公共定义域的 pdf：无论θ取何值，样本的范围是相同的。而对于一个 pdf 为$f_X(x)=I_{[\theta-1/2,\theta+1/2]}(x)$的均匀随机变量的情况则不同：$\theta$值的改变显然直接决定了可能的输出的范围。对于具有公共定义域的情况，假设未知参数取有限个值。

定理 9-2 设$f_X(x;\theta_m)$是X的 pdf，参数为θ_m，有$m=1,\cdots,M$个可能的值，则 pdf 族具有公共定义域。以下统计量是最小充分的：

$$\boldsymbol{T}=\left[\frac{f_X(\boldsymbol{x};\theta_2)}{f_X(\boldsymbol{x};\theta_1)},\cdots,\frac{f_X(\boldsymbol{x};\theta_M)}{f_X(\boldsymbol{x};\theta_1)}\right]^{\mathrm{T}} \tag{9-31}$$

证明：对充分统计量T，由因式分解可知$f_X(\boldsymbol{x};\theta)=g(t;\theta)h(\boldsymbol{x})$，其中$g(t;\theta)$只通过$t$与样本相关。因为$h(\boldsymbol{x})$不是$\theta$的函数，我们得到以下比值，对$\theta_1\neq\theta_2$，只通过$t$与样本相关：

$$\frac{f_X(\boldsymbol{x};\theta_2)}{f_X(\boldsymbol{x};\theta_1)}=\frac{g(t;\theta_2)}{g(t;\theta_1)} \tag{9-32}$$

由于$\boldsymbol{T}$包括了所有有限数目参数值的可能的比值，并且是T的函数，它必定是最小充分的。

通常，我们只需检查一个θ值的比值即可。具体需要的比值数由作为θ的函数的联合 pdf 的平滑程度决定。这个过程也可扩展运用到可数无穷个θ，但对于不可数个θ，则要复杂的多，因为要考虑测度论了。

例 9-10 考虑参数$\theta=\lambda$的指数 pdf，参数取有限个值，包括$\lambda_1=1$。由定理 9-2，则以下比值是最小充分的：

$$\frac{\prod_{n=1}^{N}\lambda\exp(-\lambda x_n)I_{[0,\infty)}(x_n)}{\prod_{n=1}^{N}\exp(-x_n)I_{[0,\infty)}(x_n)}=\frac{\lambda^N\exp\left(-\lambda\sum_{n=1}^{N}x_n\right)}{\exp\left(-\sum_{n=1}^{N}x_n\right)}=\lambda^N\exp\left(-(\lambda-1)\sum_{n=1}^{N}x_n\right) \tag{9-33}$$

它与样本均值 $T=\overline{X}$ 等价。 ◀

另一种找具有公共定义域的随机变量的最小充分统计量的方法，包括两个不同的 N 个样本集，我们不作证明的给出。

定理 9-3 设 $\boldsymbol{X}$ 是 pdf 为 $f_X(x;\theta)$ 的一个随机变量的 iid 样本，$\boldsymbol{Y}$ 为由同样分布得到的任意其他 iid 的样本集。T 是 θ 最小充分量，如果比值

$$\frac{f_X(\boldsymbol{x};\theta)}{f_X(\boldsymbol{y};\theta)}=\frac{h(\boldsymbol{x})g(t_x;\theta)}{h(\boldsymbol{y})g(t_y;\theta)} \tag{9-34}$$

当且仅当 $T_X=T_Y$ 时与 θ 独立。

T_X，T_Y 下标表示统计量中用到的样本。注意式(9-32)中参数不同，而式(9-34)中样本不同。

例 9-11 对例 9-10 的条件，我们得到：

$$\frac{\lambda^N\exp\left(-\lambda\sum_{n=1}^{N}x_n\right)}{\lambda^N\exp\left(-\lambda\sum_{n=1}^{N}y_n\right)}=\exp\left(-\lambda\sum_{n=1}^{N}x_n+\lambda\sum_{n=1}^{N}y_n\right) \tag{9-35}$$

显然，当且仅当两个求和项相等时，这个比值独立于 λ；则 $T=\sum_{n=1}^{N}X_n$ 是 $\theta=\lambda$ 的最小充分统计量。 ◀

对无公共定义域的随机变量，找最小充分统计量的方法是基于这样一个思路：由于 θ 在变化，需从 $g(t;\theta)$ 中"捕捉"充分统计量 T。图 9-3 描述了函数 $g(t;\theta)$ 与 θ 的对应关系，所有不连续的位置也都进行了标注。最小充分统计量与样本对应，样本中有不连续发生。

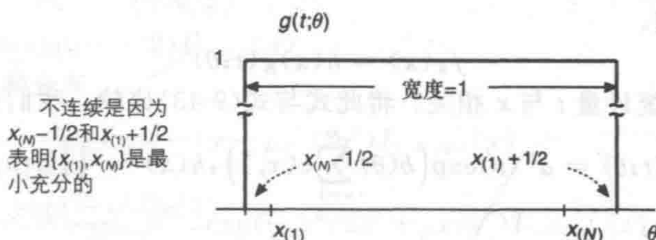


图 9-3 对 $f_X(x)=I_{[\theta-1/2,\theta+1/2]}(x)$ 的"捕捉"技术显示 $\boldsymbol{T}=[X_{(1)},X_{(N)}]^{\mathrm{T}}$ 是最小充分的。(注意到纵轴上的"断续"线是为了保持图形的紧凑)

例 9-12 对于之前的均匀分布，我们重复运用因式分解：

$$f_{\boldsymbol{X}}(\boldsymbol{x})=I_{[\theta-1/2,\theta+1/2]}(x_{(1)})I_{[\theta-1/2,\theta+1/2]}(x_{(N)}) \tag{9-36}$$

观察到指示函数可重新写成 θ 表示的形式，如下所示：

$$I_{[\theta-1/2,\theta+1/2]}(x_{(1)})=\begin{cases}1, & \theta-1/2\leqslant x_{(1)}\leqslant\theta+1/2\\ 0, & \text{其他}\end{cases} \tag{9-37}$$

尽管有可能用以上的形式来确定最小充分量，但如果其中一个边界是 $\pm\infty$，就会更容易些。这可由指示函数的以下特性看出：

$$I_{[\theta-1/2,\theta+1/2]}(x)=I_{[\theta-1/2,\infty]}(x)I_{[-\infty,\theta+1/2]}(x) \tag{9-38}$$

其中式子左边的封闭间隔是式子右边两个半封闭间隔的交集。对这个问题：

$$I_{[\theta-1/2,\theta+1/2]}(x_{(1)})I_{[\theta-1/2,\theta+1/2]}(x_{(N)})=I_{[\theta-1/2,\infty]}(x_{(1)})I_{[-\infty,\theta+1/2]}(x_{(N)}) \tag{9-39}$$

由此我们得到 $g(t;\theta)=1$，当

$$\theta-1/2\leqslant x_{(1)}\Rightarrow\theta\leqslant x_{(1)}+1/2 \tag{9-40}$$

$$x_{(N)}\leqslant\theta+1/2\Rightarrow\theta\geqslant x_{(N)}-1/2 \tag{9-41}$$

作为 θ 的函数，$g(t;\theta)$ 的非零范围在图 9-3 中说明了，不连续发生在 $x_{(N)}-1/2$ 和

$x_{(1)}+1/2$ 处。因此，$\boldsymbol{T}=[X_{(1)}, X_{(N)}]^{\mathrm{T}}$ 对于 θ 不仅仅是充分的还是最小充分的。◀

9.5 指数族

本节，我们考虑随机变量的指数族，对这种情况低维的充分统计量很容易找到。

定义(指数族) 具有标量参数 θ 的随机变量的一个**指数族**具有可表示为如下形式的 pdf：

$$f_X(x)=a(\theta)\exp(b(\theta)c(x))d(x) \tag{9-42}$$

其中 $a(\theta)$，$b(\theta)$，$c(x)$ 和 $d(x)$ 是已知函数。它被称作随机变量族是因为上式的形式在 θ 变化时是相同的。对于 N 个 iid 随机变量 $\{X_n\}$，式(9-42)变为：

$$f_{\boldsymbol{X}}(\boldsymbol{x})=a^N(\theta)\exp\Big(b(\theta)\sum_{n=1}^{N}c(x_n)\Big)\prod_{n=1}^{N}d(x_n) \tag{9-43}$$

对于 M 个参数的向量 $\boldsymbol{\theta}\triangleq[\theta_1\cdots,\theta_M]^{\mathrm{T}}$ 以及一个单一的随机变量，pdf 为：

$$f_X(x)=a(\boldsymbol{\theta})\exp\Big(\sum_{m=1}^{M}b_m(\boldsymbol{\theta})c_m(x)\Big)d(x) \tag{9-44}$$

对 N 个 iid 随机变量以及 M 个参数，式(9-44)变为：

$$f_{\boldsymbol{X}}(\boldsymbol{x})=a^N(\boldsymbol{\theta})\exp\Big(\sum_{m=1}^{M}b_m(\boldsymbol{\theta})\sum_{n=1}^{N}c_m(x_n)\Big)\prod_{n=1}^{N}d(x_n) \tag{9-45}$$

不要将随机变量的指数族与第 3 章中描述的指数分布混淆(它本身是一个指数族)。以下将证明一个指数族的 pdf 可被重置(有时是用一种不明显的方式)来得到上式指数项的形式中的一种。尽管一个 pdf 可能包括多个参数，指数形式特指我们感兴趣的标量参数 θ 或向量参数 $\boldsymbol{\theta}$；为实现估计，pdf 的其他参数假设都是已知的(固定的)。例如，对于高斯分布，我们可能只对均值 μ 感兴趣，则假设标准差 σ 是已知的。

由因式分解定理，

$$f_{\boldsymbol{X}}(\boldsymbol{x})=h(\boldsymbol{x})g(t;\theta) \tag{9-46}$$

其中 $g(t;\theta)$ 只通过统计量 t 与 $\boldsymbol{x}$ 相关。将此式与式(9-43)比较，我们得到

$$g(t;\theta)=a^N(\theta)\exp\Big(b(\theta)\sum_{n=1}^{N}c(x_n)\Big),h(\boldsymbol{x})=\prod_{n=1}^{N}d(x_n) \tag{9-47}$$

因此，充分统计量为：

$$T=\sum_{n=1}^{N}c(X_n) \tag{9-48}$$

对以下指数族的每一个随机变量都可以得到这个表达式；尽管我们只对 $N=1$ 的样本的指数形式进行说明，对 N 个样本的形式很容易由式(9-43)确定，稍后会在表 9-2 中总结。

表 9-2 指数族的最小充分统计量

统计量 $\boldsymbol{T}$	分布族(参数 $\boldsymbol{\theta}$)
$\sum_{n=1}^{N}X_n$	指数(λ)，伽马$(\lambda$，已知 $r)$ 高斯$(\mu$，已知 $\sigma)$，泊松(α) 伯努利(p)，二项式(p)，几何(p) 负二项式$(p$，已知 $N)$
$\left[\prod_{n=1}^{N}X_n,\sum_{n=1}^{N}X_n\right]^{\mathrm{T}}$	贝塔 (α,β)
$\prod_{n=1}^{N}X_n$	卡方
$\left[\sum_{n=1}^{N}X_n,\sum_{n=1}^{N}X_n^2\right]^{\mathrm{T}}$	高斯 (μ,σ)
$\sum_{n=1}^{N}\lvert X_n\rvert$	拉普拉斯$(\alpha$，已知 $\mu=0)$

● $\theta=\lambda$ 的指数分布：

$$f_X(x)=\lambda\exp(-\lambda x)I_{[0,\infty)}(x) \tag{9-49}$$

$$\Rightarrow a(\theta)=\lambda,\quad b(\theta)=-\lambda,\quad c(x)=x,\quad d(x)=I_{[0,\infty)}(x) \tag{9-50}$$

● $\theta=\lambda$ 的伽马分布(已知 r)：

$$f_X(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}\exp(-\lambda x)I_{[0,\infty)}(x) \tag{9-51}$$

$$\Rightarrow a(\theta)=\frac{\lambda^r}{\Gamma(r)},\quad b(\theta)=-\lambda,\quad c(x)=x,\quad d(x)=x^{r-1}I_{[0,\infty)}(x) \tag{9-52}$$

● $\theta=\{\alpha,\beta\}$ 的贝塔：

$$\begin{aligned}f_X(x)&=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}I_{[0,1]}(x)\\&=\frac{1}{B(\alpha,\beta)}[\exp((\alpha-1)\ln(x)+(\beta-1)\ln(1-x)]I_{[0,1]}(x)\end{aligned} \tag{9-53}$$

$$\Rightarrow a(\boldsymbol{\theta})=\frac{1}{B(\alpha,\beta)},\quad b_1(\boldsymbol{\theta})=\alpha-1,\quad b_2(\boldsymbol{\theta})=\beta-1 \tag{9-54}$$

$$c_1(x)=\ln(x),c_2(x)=\ln(1-x),d(x)=I_{[0,1])}(x) \tag{9-55}$$

● $\theta=N$ 的卡方分布：

$$\begin{aligned}f_X(x)&=\frac{x^{N/2-1}}{2^{N/2}\Gamma(N/2)}\exp(-x/2)I_{[0,\infty)}(x)\\&=\frac{\exp((N/2-1)\ln(x))}{2^{N/2}\Gamma(N/2)}\exp(-x/2)I_{[0,\infty)}(x)\end{aligned} \tag{9-56}$$

$$\Rightarrow a(\theta)=\frac{2^{-N/2}}{\Gamma(N/2)},\quad b(\theta)=N/2-1,\quad c(x)=\ln(x) \tag{9-57}$$

$$d(x)=\exp(-x/2)I_{[0,\infty)}(x) \tag{9-58}$$

● $\boldsymbol{\theta}=\{\mu,\sigma\}$ 的高斯分布：

$$\begin{aligned}f_X(x)&=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/2\sigma^2)I_{(-\infty,\infty)}(x)\\&=\frac{\exp(-\mu^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}\exp(-x^2/2\sigma^2+\mu x/\sigma^2)I_{(-\infty,\infty)}(x)\end{aligned} \tag{9-59}$$

$$\Rightarrow a(\boldsymbol{\theta})=\frac{\exp(-\mu^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}},\quad b_1(\boldsymbol{\theta})=-\frac{1}{2\sigma^2},\quad b_2(\boldsymbol{\theta})=\frac{\mu}{\sigma^2} \tag{9-60}$$

$$c_1(x)=x^2,\quad c_2(x)=x,\quad d(x)=I_{(-\infty,\infty)}(x) \tag{9-61}$$

● $\theta=\alpha$ 的瑞利分布：

$$f_X(x)=(x/\alpha^2)\exp(-x^2/2\alpha^2)I_{[0,\infty)}(x) \tag{9-62}$$

$$\Rightarrow a(\theta)=1/\alpha^2,b(\theta)=-1/2\alpha^2,\quad c(x)=x^2,\quad d(x)=xI_{[0,\infty)}(x) \tag{9-63}$$

● $\theta=\alpha$ 的拉普拉斯分布(已知 $\mu=0$)：

$$f_X(x)=(1/2\alpha)\exp(-|x|/\alpha)I_{(-\infty,\infty)}(x) \tag{9-64}$$

$$\Rightarrow a(\theta)=1/2\alpha,\quad b(\theta)=-1/\alpha,\quad c(x)=|x|,\quad d(x)=I_{(-\infty,\infty)}(x) \tag{9-65}$$

● $\theta=\alpha$ 的泊松分布：

$$\begin{aligned}p_X[x]&=(1/x!)\exp(-\alpha)\alpha^x I_{z^+}[x]\\&=(1/x!)\exp(-\alpha)\exp(x\ln(\alpha))I_{z^+}[x]\end{aligned} \tag{9-66}$$

$$\Rightarrow a(\theta)=\exp(-\alpha),\quad b(\theta)=\ln(\alpha),\quad c(x)=x,\quad d(x)=(1/x!)I_{z^+}[x] \tag{9-67}$$

● $\theta=p$ 的几何分布：

$$p_X[x]=pq^x I_{z^+}[x]=p\exp(x\ln(q))I_{z^+}[x] \tag{9-68}$$

$$\Rightarrow a(\theta)=p,\quad b(\theta)=\ln(q),\quad c(x)=x,\quad d(x)=I_{z^+}[x] \tag{9-69}$$

● $\theta=p$ 的二项式分布(已知 N)：

$$p_X[x]=\binom{N}{x}p^x q^{N-x}I_{\{0,1,\cdots,N\}}[x]=\binom{N}{x}q^N\exp(x\ln(p/q))I_{\{0,1,\cdots,N\}}[x] \tag{9-70}$$

$$\Rightarrow a(\theta)=q^N,\quad b(\theta)=\ln(p/q),\quad c(x)=x,\quad d(x)=\binom{N}{x}I_{\{0,1,\cdots,N\}}[x] \tag{9-71}$$

即使二项式概率质量函数(pmf)有两个参数 p 和 q，它仍是一个单参数分布，因为 $q\triangleq 1-p$。稍后例 9-14 中讨论的多项式分布也有类似的约束，我们写的是 M 个参数，但实际只有 $M-1$ 个参数。

- $\theta=p$ 的伯努利分布($N=1$ 的二项式)：

$$p_X[x]=p^x q^{1-x}I_{\{0,1\}}[x]=q\exp(x\ln(p/q))I_{\{0,1\}}[x] \tag{9-72}$$

$$\Rightarrow a(\theta)=q,\quad b(\theta)=\ln(p/q),\quad c(x)=x,\quad d(x)=I_{\{0,1\}}[x] \tag{9-73}$$

- $\theta=p$ 的负二项式分布(已知 N)：

$$p_X[x]=\binom{x+N-1}{x}p^N q^x I_{Z^+}[x]=\binom{x+N-1}{x}p^N\exp(x\ln(q))I_{Z^+}[x], \tag{9-74}$$

$$\Rightarrow a(\theta)=p^N,\quad b(\theta)=\ln(q),\quad c(x)=x,\quad d(x)=\binom{x+N-1}{x}I_{Z^+}[x] \tag{9-75}$$

由于指数族具有与式(9-47)的因式分解定理匹配的形式，对每个上文中提到族的最小充分统计量非常容易确认。表 9-2 中进行了总结。

例 9-13 贝塔分布的联合 pdf 为

$$f_{\boldsymbol{X}}(\boldsymbol{x})=\frac{1}{B^N(\alpha,\beta)}\exp\Big((\alpha-1)\sum_{n=1}^{N}\ln(x_n)+(\beta-1)\sum_{n=1}^{N}\ln(1-x_n)\Big)\cdot\prod_{n=1}^{N}I_{(0,1)}(x_n) \tag{9-76}$$

由于

$$\sum_{n=1}^{N}\ln(x_n)=\ln\Big(\prod_{n=1}^{N}x_n\Big),\quad \sum_{n=1}^{N}\ln(1-x_n)=\ln\Big(\prod_{n=1}^{N}(1-x_n)\Big) \tag{9-77}$$

则 $\boldsymbol{\theta}=[\alpha,\beta]^{\mathrm{T}}$ 的最小充分统计量为

$$\boldsymbol{T}=\Big[\prod_{n=1}^{N}X_n,\prod_{n=1}^{N}(1-X_n)\Big]^{\mathrm{T}} \tag{9-78}$$ ◀

例 9-14 多项式的 pdf 要写成指数族的形式在某种程度上要更复杂些，因为参数并不都是独立的，因此它会被作为一个单独的例子。设 $\{X_1,\cdots,X_M\}$ 为参数(概率)为 $\{p_1,\cdots,p_M\}$ 的 M 个样本。对已知的 M，得到 $\sum_{l=1}^{M}x_l=N$，多项式的 pdf 为

$$\begin{aligned}f_{X_1,\cdots,X_M}(x_1,\cdots,x_M)&=N!\prod_{m=1}^{M}\frac{p_m^{x_m}}{x_m!}I_{\{0,1,\cdots,N\}}(x_m)I_{\{N\}}\Big(\sum_{l=1}^{M}x_l\Big)\\&=N!\exp\Big(\sum_{m=1}^{M}x_m\ln(p_m)\Big)\\&\quad\cdot\prod_{m=1}^{M}\frac{1}{x_m!}I_{\{0,1,\cdots,N\}}(x_m)I_{\{N\}}\Big(\sum_{l=1}^{M}x_l\Big)\end{aligned} \tag{9-79}$$

由于 $\sum_{m=1}^{M}p_m=1$，所以只有 $M-1$ 个独立的参数；因此，可以很方便地将参数向量写为 $\boldsymbol{\theta}=(p_2/p_1,\cdots,p_M/p_1)$(回顾一下比值 p/q 曾用在二项式分布中，它是多项式分布在 $M=2$ 时的特例)。加或减 $\sum_{m=2}^{M}x_m\ln(p_1)$ 到 $\exp(\cdot)$的指数项中，代入约束 $\sum_{m=1}^{M}x_m=N$ 则得到：

$$\exp\Big(N\ln(p_1)+\sum_{m=2}^{M}x_m\ln(p_m/p_1)\Big)=\exp\Big(-N\ln(1+\sum_{m=2}^{M}p_m/p_1)\Big)\cdot\exp\Big(\sum_{m=2}^{M}x_m\ln(p_m/p_1)\Big) \quad (9\text{-}80)$$

其中我们已做了替换

$$1/p_1=\sum_{m=1}^{M}p_m/p_1=1+\sum_{m=2}^{M}p_m/p_1 \quad (9\text{-}81)$$

在等号右边的第一个指数项中将 $\ln(p_1)$ 写成比值 $\{p_2/p_1,\cdots,p_M/p_1\}$ 的形式。指数族形式具有分量：

$$a(\boldsymbol{\theta})=N!\exp\Big(-N\ln\Big(1+\sum_{m=2}^{M}p_m/p_1\Big)\Big),b_m(\boldsymbol{\theta})=\ln(p_m/p_1) \quad (9\text{-}82)$$

$$c_m(x_1,\cdots,x_M)=x_m,d(x_1,\cdots,x_M)=\prod_{m=1}^{M}\frac{1}{x_m!}I_{\{0,1,\cdots,N\}}(x_m)I_{\{N\}}\Big(\sum_{l=1}^{M}x_l\Big) \quad (9\text{-}83)$$

9.6 位置-比例族

随机变量的位置-比例族的 pdf 在它的位置或比例参数变化时仍保持它的基本形状。

定义（位置-比例族）　设 X 为一个随机变量，cdf 为 $F_X(x)$。通过变换 $Y=X+u$ 可得到一个**位置族**，其中 $u\in\mathcal{R}$。它的分布具有以下形式：

$$F_Y(y)=F_X(y-u) \quad (9\text{-}84)$$

通过变换 $Y=vX$ 可得到一个**比例族**，其中 $v>0$。它的分布具有如下形式：

$$F_Y(y)=F_X(y/v) \quad (9\text{-}85)$$

将两个变换组合为 $Y=u+vX$，得到一个**位置-比例族**，它的分布形式为：

$$F_Y(y)=F_X((y-u)/v) \quad (9\text{-}86)$$

pdf 为

$$f_Y(y)=(1/v)f_X((y-u)/v) \quad (9\text{-}87)$$

其中 $f_X(x)$ 是在变换成 Y 之前 X 的 pdf。

表 9-3 提供了几个位置－比例族的最小充分统计量。位置-比例族有两个参数，包括以下的随机变量。对每一种情况，我们都可以通过变换 $Y=u+vX$ 来分析位置-比例族的特性，并且证明 $f_Y(y)=(1/v)f_X((y-u)/v)$ 与 $f_X(x)$ 具有相同的形式。

表 9-3　位置-比例族的最小充分统计量

统计量 T	分布族（参数 θ）
$[X_{(1)},\cdots,X_{(N)}]^{\mathrm{T}}$	柯西$\{c$，已知 $\alpha\}$，拉普拉斯$\{\mu$，已知 $\alpha\}$，符号逻辑$\{\mu$，已知 $\alpha\}$
$X_{(1)}$	偏移指数$\{c$，已知 $\lambda\}$，均匀$\{a$，已知 $b\}$
$\Big[X_{(1)},\sum_{n=1}^{N}X_n\Big]^{\mathrm{T}}$	偏移指数$\{c,\lambda\}$
$[X_{(1)},X_{(N)}]^{\mathrm{T}}$	均匀$\{a,b\}$
$X_{(N)}$	均匀$\{b$，已知 $a\}$
$\Big[\sum_{n=1}^{N}X_n\sum_{n=1}^{N}X_n^2\Big]^{\mathrm{T}}$	高斯$\{\mu,\sigma\}$

- $\boldsymbol{\theta}=\{\mu,\sigma\}$ 的高斯分布：

$$f_Y(y)=\frac{1}{\sqrt{2\pi v^2\sigma^2}}\exp(-((y-u)/v-\mu)^2/2\sigma^2)I_{(-\infty,\infty)}(y)$$

$$=\frac{1}{\sqrt{2\pi(v\sigma)^2}}\exp(-(y-u-v\mu)^2/2(v\sigma)^2)I_{(-\infty,\infty)}(y) \quad (9\text{-}88)$$

⇒参数为$\{u+v\mu, v\sigma\}$的高斯分布。

- $\boldsymbol{\theta}=\{a, b\}$的均匀分布：

$$f_Y(y)=\frac{1}{v(b-a)}I_{[a,b]}((y-u)/v)=\frac{1}{vb-va}I_{[u+va,u+vb]}(y) \tag{9-89}$$

⇒参数为$\{u+va, u+vb\}$的均匀分布。

- $\boldsymbol{\theta}=\{c, \alpha\}$的柯西分布：

$$f_Y(y)=\frac{\alpha/v\pi}{((y-u)/v-c)^2+\alpha^2}I_{(-\infty,\infty)}(y)=\frac{v\alpha/\pi}{(y-u-vc)^2+(v\alpha)^2}I_{(-\infty,\infty)}(y) \tag{9-90}$$

⇒参数为$\{u+vc, v\alpha\}$的柯西分布。

- $\boldsymbol{\theta}=\{\mu, \alpha\}$的拉普拉斯分布：

$$\begin{aligned}f_Y(y)&=(1/2v\alpha)\exp(-|(y-u)/v-\mu|/\alpha)I_{(-\infty,\infty)}(y)\\&=(1/2v\alpha)\exp(-|y-u-v\mu|/v\alpha)I_{(-\infty,\infty)}(y)\end{aligned} \tag{9-91}$$

⇒参数为$\{u+vc, v\alpha\}$的拉普拉斯分布。

- $\boldsymbol{\theta}=\{\mu, \alpha\}$的 Logistic 分布

$$\begin{aligned}f_Y(y)&=\frac{\exp(-((y-u)/v-\mu)/\alpha)}{v\alpha(1+\exp(-((y-u)/v-\mu)/\alpha))}I_{(-\infty,\infty)}(y)\\&=\frac{\exp(-(y-u-v\mu)/v\alpha)}{v\alpha(1+\exp(-(y-u-v\mu)/v\alpha))}I_{(-\infty,\infty)}(y)\end{aligned} \tag{9-92}$$

⇒参数为$\{u+v\mu, v\alpha\}$的 Logistic 分布。

- $\boldsymbol{\theta}=\{c, \lambda\}$的偏移指数分布：

$$\begin{aligned}f_Y(y)&=(\lambda/v)\exp(-\lambda((y-u)/v-c))I_{[c,\infty)}((y-u)/v)\\&=(\lambda/v)\exp(-(\lambda/v)(y-u-vc))I_{[u+vc,\infty)}(y)\end{aligned} \tag{9-93}$$

⇒参数为$\{u+vc, \lambda/v\}$的偏移指数。

- $\boldsymbol{\theta}=\{a, b\}$，$N\triangleq b-a+1$ 的离散均匀分布

$$p_Y[y]=(1/Nv)I_{\{a,\cdots,b\}}[(y-u)/v]=(1/Nv)I_{\{u+va,\cdots,u+vb\}}[y] \tag{9-94}$$

⇒$N=(u+vb)-(u+va)+1=v(b-a)+1$，参数$\{u+va, u+vb\}$的离散均匀分布（假设选择合适的$u$和$v$，使参数为整数）。

$\boldsymbol{\theta}=\{\mu, \sigma\}$的高斯随机变量既是指数族又是位置-比例族。但是，这是例外，因为绝大多数位置-比例族不能被写成式(9-42)的指数形式。例如，通过扩展指数 pdf 包含了移位参数c后，它就不再是一个指数族了。

例 9-15 考虑N个$\alpha=1$的柯西分布的 iid 样本，得到联合 pdf：

$$f_{\boldsymbol{X}}(\boldsymbol{x})=\frac{1}{\pi^N}\prod_{n=1}^{N}\frac{1}{(x_n-c)^2+1}I_{(-\infty,\infty)}(x_n) \tag{9-95}$$

看起来不可能通过简化这个表达式来得到$\theta=c$的维度小于N的充分统计量。经常会出现这种结果(但也不总是这样)，顺序统计量是位置-比例族的位置参数的最小充分统计量，就如我们现在证明的柯西分布的情况。考虑另一个样本集$\boldsymbol{y}$，得到式(9-34)的比例关系：

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x};c)}{f_{\boldsymbol{X}}(\boldsymbol{y};c)}=\prod_{n=1}^{N}\frac{(y_n-c)^2+1}{(x_n-c)^2+1}=\prod_{n=1}^{N}\frac{c^2-2cy_n+y_{n+1}^2}{c^2-2cx_n+x_{n+1}^2} \tag{9-96}$$

对乘积项中的每一个二次项，设分子的根为$c=y_n\pm\mathrm{j}$，分母的根为$c=x_n\pm\mathrm{j}$，其中$\mathrm{j}=\sqrt{-1}$，则：

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x};c)}{f_{\boldsymbol{X}}(\boldsymbol{y};c)}=\prod_{n=1}^{N}\frac{(c-y_n-\mathrm{j})(c-y_n+\mathrm{j})}{(c-x_n-\mathrm{j})(c-x_n+\mathrm{j})} \tag{9-97}$$

如果忽略式子相乘的顺序，乘积项是相同的，所以使比例式与c独立的唯一方法是分

子式的根与分母式的根排序相同。这意味着分子和分母必须有相同的顺序统计量，因此 $\boldsymbol{T}=[X_{(1)},\cdots,X_{(N)}]^{\mathrm{T}}$ 是柯西分布的 c 的最小充分量。习题 9-8 中证明了当均值 μ 是待估参量时，顺序统计量同样也是拉普拉斯分布的最小充分量。◀

9.7　完备统计量

尽管 $\boldsymbol{T}=[X_{(1)},\ X_{(N)}]^{\mathrm{T}}$ 是例 9-12 中均匀分布的 θ 的最小充分量，它们所包含的 θ 的信息与初始样本一样多，但他们也包括一些与估计 θ 无关的信息。这些无关信息被称做从属项，一个最小充分统计量的任何从属分量被称为从属统计量。一个不含从属分量的最小充分统计量被称为一个完备最小充分统计量。我们给出一个从属统计量的两个定义。

定义(从属统计量)　如果统计量 V 的分布独立于 θ，则其为从属的。

从属统计量不包括参数 θ 的任何信息。这与充分统计量的特性有本质上的区别，充分统计量包含的关于 θ 的信息与初始样本一样。

定义(一阶从属统计量)　统计量 V 是一阶从属的，如果 $\varepsilon[V]$ 与 θ 独立。

一个从属统计量也是一阶从属的：如果它的分布独立于 θ，则它的期望值不依赖于 θ。

例 9-16　对均匀 pdf $f_X(x)=I_{[\theta-1/2,\theta+1/2]}(x)$，观察到统计量 $V=X_{(N)}-X_{(1)}$ 是一阶从属的，因为它的均值与 θ 无关：

$$\mathcal{E}[X_{(N)}-X_{(1)}]=1-\frac{2}{N+1}=\frac{N-1}{N+1} \tag{9-98}$$

其中用到了式(9-29)和式(9-30)。我们已能找到一个最小充分统计量的函数，它是一阶从属的。注意到(ⅰ)V 与 pdf 的宽度相关(它独立于 θ)；(ⅱ)上式的期望值渐进趋于 1。

接下来的定义指的是这样一类统计量，它不包括 θ 的从属信息。

定义(完备统计量)　统计量 T 是完备的，如果对每个函数 $c(\cdot)$有：

$$\mathcal{E}[c(T)]=0(\text{对所有的}\ \theta) \tag{9-99}$$

意味着：

$$P(c(T)=0)=1(\text{对所有的}\ \theta) \tag{9-100}$$

一个完备统计量有可能不包括 θ 的信息。因此，我们感兴趣的是一个统计量，它即是充分的，又是完备的，这样它就不会丢失任何关于 θ 的信息，也不会有无关分量。接下来的定理将证明一个完备充分统计量没有从属分量。

定理 9-4(Basu)　设 T 为 θ 的一个完备充分统计量。任意从属统计量 V 与 T 独立。

证明：我们要证明 $P(V\leqslant v|T=t)=P(V\leqslant v)$。由于 V 是从属的，$P(V\leqslant v)=F_V(v)$ 与 θ 无关。定义以下条件概率：

$$g(t)\triangleq P(V\leqslant v|T=t) \tag{9-101}$$

它是 t 的函数，得到它的均值为：

$$\mathcal{E}[g(T)]=\int_{-\infty}^{\infty}P(V\leqslant v|T=t)f_T(t)\mathrm{d}t=P(V\leqslant v) \tag{9-102}$$

将 $F_V(v)$代入式(9-102)左边得到：

$$\mathcal{E}[g(T)-F_V(v)]=0\Rightarrow P(g(T)-F_V(v)=0)=1\Rightarrow P(g(T)=F_V(v))=1 \tag{9-103}$$

由于 T 是完备的。完备性定义中的函数此时对应为 $c(T)=g(T)-F_V(v)$。最后的结果意味着对于所有 t 和所有 θ，$P(V\leqslant v|T=t)=F_V(v)$的概率为 1，证明完毕。

例 9-17　设 $\{X_1,\cdots,X_N\}$ 为 iid 高斯随机变量，参数为 $\{\mu,\sigma=1\}$。习题 9-10 中证明了样本均值 $T=\overline{X}$ 是 μ 的一个完备充分统计量。习题 4-35 证明了$(N-1)S^2/\sigma^2$ 具有卡方分布，参数为 $N-1$，其中 S^2 为样本方差。由于 $V=S^2$ 的分布不依赖于 μ，它是从属统计量，由 Basu 定理，对高斯分布而言，样本均值和方差是独立的。◀

例 9-18 考虑一个在$[a,\theta]$上均匀分布的随机变量，θ是独立参数，a已知，$\theta>a$。对N个 iid 样本，联合 pdf 为：

$$f_{\mathbf{X}}(\mathbf{x})=\frac{1}{(\theta-a)^N}I_{[a,\theta]}(x_{(N)}) \tag{9-104}$$

从中可看出$X_{(N)}$对θ是最小充分的，因为它是一维的。而且，它是一个完备充分统计量，接下来将证明这一点。由定义，我们先计算以下期望值：

$$\mathcal{E}[c(X_{(N)})]=N\int_{-\infty}^{\infty}c(x)F_X^{N-1}(x)f_X(x)\mathrm{d}x \tag{9-105}$$

对均匀 pdf 有：

$$F_X(x)=\frac{1}{\theta-a}\int_{-\infty}^{x}I_{[a,\theta]}(x)\mathrm{d}x=\frac{x-a}{\theta-a}I_{[a,\theta]}(x) \tag{9-106}$$

将 pdf 和 cdf 代入得到：

$$\mathcal{E}[c(X_{(N)})]=\frac{N}{(\theta-a)^N}\int_a^{\theta}c(x)(x-a)^{N-1}\mathrm{d}x \tag{9-107}$$

将此结果置 0，消去积分项前的常数，对θ微分得到：

$$c(\theta)(\theta-a)^{N-1}=0 \tag{9-108}$$

它对所有$\theta>a$的值都必须成立，这意味着几乎对所有θ值，函数依概率 1 满足$c(\theta)=0$。则对任意有限值a，$X_{(N)}$是θ的一个完备最小充分统计量。◀

例 9-19 设$\{X_1,\cdots,X_N\}$是 iid 伯努利分布，参数为p：

$$P(\mathbf{X}=\mathbf{x})=p^{\sum_{n=1}^{N}x_n}q^{1-\sum_{n=1}^{N}x_n}\prod_{n=1}^{N}I_{\{0,1\}}[x_n] \tag{9-109}$$

由表 9-2，我们知道$T=\sum_{n=1}^{N}X_n$是p的最小充分统计量。T的分布为二项式，参数为$\{N,p\}$，则

$$\mathcal{E}[c(T)]=\sum_{n=0}^{N}c(n)\binom{N}{n}p^nq^{N-n}=c(0)q^N+c(1)Npq^{N-1}+\cdots+c(N)p^N \tag{9-110}$$

因为p和q的指数项在求和式中的各项都不同，上式对所有p都为 0 的唯一办法是对所有n有$c(n)=0$。因此，T也是完备的。◀

9.8 节的目标是要用一个充分统计量的完备性得到一个最优估计器，它不仅要是无偏的，还是所有无偏估计器中方差最小的。它被称为一致最小方差无偏(UMVU)估计器。

定义(UMVU 估计器)　基于N个 iid 样本$\{X_1,\cdots,X_N\}$的参数θ的 UMVU 估计器T_{UMVU}具有如下特性：

- 无偏性。$\mathcal{E}[T_{\mathrm{UMVU}}]=\theta$
- 最小方差。对任意其他θ的无偏估计器T满足$\mathrm{var}[T_{\mathrm{UMVU}}]\leqslant\mathrm{var}[T]$。

在开始介绍如何运用完备性来产生这样一个估计器之前，我们先介绍一个生成较低方差的估计器的方法。

9.8　拉奥-布莱克维尔定理

拉奥-布莱克维尔(RB)定理描述了一种改进相对较差的θ的估计器，它通常是直接去找使其具有更低方差的方法。为了证明这个定理，我们需要附录 F 中的杰森不等式：

$$g(\mathcal{E}[X])\leqslant\mathcal{E}[g(X)] \tag{9-111}$$

其中X是一个定义在开区间上的随机变量，cdf 为$F_X(x)$，$g(x)$是一个凸函数(见附录 B)。尽管拉奥-布莱克维尔定理可用于任意凸函数，就如用在杰森不等式中，但我们仍给出其用于估计器方差的定理。

定理 9-5(拉奥-布莱克维尔定理)　设$\{T_1,\cdots,T_P\}$是一个统计量集合，它们对参数θ的函数$c(\theta)$是联合充分的。定义另一个统计量T，它是$c(\theta)$的无偏估计器。则统计量

$T_{\mathrm{RB}} \triangleq \mathcal{E}[T \mid T_1, \cdots, T_p]$ 具有如下特性。

- $\mathcal{E}[T_{\mathrm{RB}}]$是 $\{T_1, \cdots, T_p\}$ 的一个函数，因此是充分的。
- $\mathcal{E}[T_{\mathrm{RB}}] = c(\theta)$，因此 T_{RB}是 $c(\theta)$的无偏估计器。
- 对所有 θ 有 $\mathrm{var}[T_{RB}] \leqslant \mathrm{var}[T]$。
- 对某些 θ，$\mathrm{var}[T_{RB}] < \mathrm{var}[T]$，除了当 T_{RB}和 T 依概率 1 相等。

证明：前两个性质由条件期望的定义得到。第三个和第四个性质由杰森不等式推导得到，只要令 $g(x) = [x - c(\theta)]^2$，则

$$(\mathcal{E}[T \mid T_1 = t_1, \cdots, T_p = t_P] - c(\theta))^2 < \mathcal{E}[(T - c(\theta))^2 \mid T_1 = t_1, \cdots, T_p = t_P] \tag{9-112}$$

式(9-112)左边包括了写成特定值 $\{t_1, \cdots, t_P\}$ 函数的 T_{RB}，式(9-112)右边是对应特定条件 $\{t_1, \cdots, t_P\}$ 的初始无偏估计器 T 的条件方差。两边同时在条件作用的变量上取期望值，得到：

$$\begin{aligned}\mathcal{E}[(\mathcal{E}[T \mid T_1 = t_1, \cdots, T_p = t_P] - c(\theta))^2] &= \mathcal{E}[(T_{\mathrm{RB}} - c(\theta))^2] = \mathrm{var}[T_{\mathrm{RB}}] \\ &< \mathcal{E}[\mathcal{E}[(T - c(\theta))^2 \mid T_1 = t_1, \cdots, T_p = t_P]] \\ &= \mathrm{var}[T]\end{aligned} \tag{9-113}$$

不等式的左边是 T_{RB}的方差，右边因为嵌套的期望值(见第 5 章)简化为 T 的方差。正如在杰森不等式证明(见附录)中所述，因为方差是严格凸的，上式是严格不等的，除非依概率 1 有 $\mathcal{E}[T \mid T_1, \cdots, T_p] = T_{\mathrm{RE}}$ 成立。

例 9-20 设 $\{X_1, \cdots, X_N\}$ 是 iid 指数随机变量，参数 $\theta = \lambda$。假设我们对一个估计器感兴趣，它估计的是：

$$c(\lambda) = p(X > x_0) = 1 - F_X(x_0) = \exp(-\lambda x_0) \tag{9-114}$$

其中 $x_0 > 0$。要采用拉奥-布莱克维尔定理需要两个分量：(ⅰ)一个对 λ 充分的统计量；(ⅱ)一个 $c(\lambda)$的无偏估计器，它不必是充分的。由例 9-4，我们得到 $T_1 = \sum_{n=1}^{N} X_n$ 对 λ 是充分的。考虑以下估计器 $T = I_{(x_0, \infty)}(X_1)$，由指示函数的特性得到它是一个 $c(\lambda)$的无偏估计器：

$$\mathcal{E}[T] = \mathcal{E}[I_{(x_0, \infty)}(X_1)] = P(X_1 > x_0) \tag{9-115}$$

直观感觉 T 是一个较差的估计器(尽管它是无偏的)，因为它只依赖于一个样本，而我们知道基于所有 N 个样本的 T_1 是充分的。为了找到 $T_{\mathrm{RB}} = \mathcal{E}[T \mid T_1]$，利用$\{X_n\}$是独立的特性推导得到条件 pdf $f_{X_1 \mid T_1}(x_1 \mid t_1)$如下：

$$\begin{aligned}P(X_1 \leqslant x_1, T_1 \leqslant t_1) &= P(X_1 \leqslant x_1, T_1 - X_1 \leqslant t_1 - x_1) \\ &= P(X_1 \leqslant x_1) P(T_1 - X_1 \leqslant t_1 - x_1) \\ &= P(X_1 \leqslant x_1) P\left(\sum_{n=2}^{N} X_n \leqslant t_1 - x_1\right)\end{aligned} \tag{9-116}$$

其中 $t_1 \geqslant x_1$。当随机变量中有一个是独立随机变量的和时常用这个技术。将这个表达式微分，得到需要的联合 pdf：

$$f_{X_1, T_1}(x_1, t_1) = f_{X_1}(x_1) I_{[0, \infty)}(x_1) f_{T_2}(t_1 - x_1) I_{[x_1, \infty)}(t_1) \tag{9-117}$$

其中 $T_2 \triangleq T_1 - X_1$ 为埃尔朗分布，参数为$\{\lambda, r = N - 1\}$，因为它是 $N-1$ 个 iid 指数随机变量的和。因此

$$\begin{aligned}f_{X_1, T_1}(x_1, t_1) &= \lambda \exp(-\lambda x_1) I_{[0, \infty)}(x_1) \frac{\lambda^{N-1}}{(N-2)!}(t_1 - x_1)^{N-2} \exp(-\lambda(t_1 - x_1)) I_{[x_1, \infty)}(t_1) \\ &= \frac{\lambda^N}{(N-2)!}(t_1 - x_1)^{N-2} \exp(-\lambda t_1) I_{[0, \infty)}(x_1) I_{[x_1, \infty)}(t_1)\end{aligned} \tag{9-118}$$

T_1 的边缘 pdf 可通过对 x_1 在 $[0, t_1]$ 上积分得到：

$$f_{T_1}(t_1) = \frac{\lambda^N}{(N-1)!} t_1^{N-1} \exp(-\lambda t_1) I_{[0,\infty)}(t_1) \tag{9-119}$$

由比值 $f_{X_1,T_1}(x_1, t_1)/f_{T_1}(t_1)$ 得到条件 pdf：

$$f_{X_1|T_1}(x_1|t_1) = (N-1)[(t_1-x_1)^{N-2}/t_1^{N-1}] I_{[0,\infty)}(x_1) I_{[x_1,\infty)}(t_1) \tag{9-120}$$

则需要的期望值为：

$$\mathcal{E}[I_{(x_0,\infty)}(X_1)|T_1=t_1] = (N-1)\int_{x_0}^{t_1}[(t_1-x_1)^{N-2}/t_1^{N-1}]\mathrm{d}x_1 = \left(\frac{t_1-x_0}{t_1}\right)^{N-1} I_{[x_0,\infty)}(t_1) \tag{9-121}$$

由此得到以下的改进的对 $c(\lambda)=\exp(-\lambda x_0)$ 的估计器：

$$T_{\mathrm{RB}} = \left[\frac{\sum_{n=1}^{N} X_n - x_0}{\sum_{n=1}^{N} X_n}\right]^{N-1} = \left[1 - \frac{x_0}{\sum_{n=1}^{N} X_n}\right]^{N-1} \tag{9-122}$$ ◀

9.9 莱赫曼-斯爵非定理

9.8 节中描述的变换并不必然产生 UMVU 估计器。它只保证了改进的估计器的方差不会超过初始统计量 T 的方差。本节的定理将描述如何得到 UMVU 估计器。我们首先给出一个关于估计器的唯一性的结论。

定理 9-6 设 T 为 θ 的一个完备充分统计量。如果存在一个 $c(\theta)$ 的无偏估计 $g(T)$，则估计器是唯一的。

证明： 定理是用反证法来证明的。考虑另一个无偏估计器 $h(T)$，$\mathcal{E}[h(T)]=c(\theta)$，它依概率与 $g(T)$ 不同：$P(h(T)-g(T)\neq 0)>0$。由于两个估计器都是无偏的：

$$\mathcal{E}[h(T)-g(T)] = 0 \tag{9-123}$$

这意味着 $h(T)-g(T)=0$，因为 T 是完备的。这与假设 $P(h(T)-g(T)\neq 0)>0$ 矛盾，所以 $h(T)$ 一定与 $g(T)$ 相同，这意味着 $g(T)$ 是唯一的。

定理 9-7 给出一个估计器的以下特性：充分性、完备性、无偏性和最小方差。

定理 9-7（莱赫曼-斯爵非） 设 T 是 θ 的一个完备充分统计量。如果 $\mathcal{E}[g(T)]=c(\theta)$，其中 $g(\cdot)$ 和 $c(\cdot)$ 是实函数，则 $g(T)$ 是 $c(\theta)$ 的 UMVU 估计器。

证明： 设 $h(\boldsymbol{X})$ 是 $c(\theta)$ 的任意无偏估计器。由定理 9-6，我们得到：

$$\mathcal{E}[h(\boldsymbol{X})|T] = g(T) \tag{9-124}$$

是 $c(\theta)$ 唯一的无偏估计器。而且，由拉奥-布莱克维尔定理，$g(T)$ 的方差不可能超过 $h(\boldsymbol{X})$ 的方差，因此 $g(T)$ 是 UMVU 估计器。

这个定理以及它的证明提供了两种不同的推导 UMVU 估计器的方法：

- 无偏性。找到 θ 的一个完备充分统计量 T，解方程 $\mathcal{E}[g(T)]=c(\theta)$，得到函数 $g(\cdot)$。回顾一个估计器仅是样本的一个特定函数。
- 条件作用。找到 θ 的一个完备充分统计量 T_1 以及 $c(\theta)$ 的一个无偏估计器 T。计算条件期望 $\mathcal{E}[T|T_1]$，它也是样本的一个特定函数。

这两种方法都需要一个 θ 的完备充分统计量。第一种方法的目标是处理 $\mathcal{E}[g(T)]$ 的方程来找到 $g(\cdot)$ 的表达式。第二种方法与拉奥-布莱克维尔定理描述的基本相同，除了此处的充分统计量 T_1 必须是完备的；由于条件期望是 T_1 的函数，由定理 9-6，它提供了唯一的 UMVU 估计器。

例 9-21 设 $\{X_n\}$ 是 iid 伯努利随机变量，参数 $\theta=p$，$T=\sum_{n=1}^{N} X_n$ 是它的一个完备充分统计量（见例 9-19）。用上文中第一种方法，我们得到：

$$(1/N)\,\mathcal{E}[T] = p \tag{9-125}$$

因此 $g(T)=\overline{X}$ 是 p 的 UMVU 估计器。第二种方法要麻烦些，但用在这里来证明完备性。注意到$\{X_n\}$中任意项可用于构造一个无偏估计器，因为$\mathcal{E}[X_n]=p$。为了计算期望值$\mathcal{E}[X_n|T]$，我们需要以下的条件概率：

$$P(X_n=x_n|T=t)=\frac{P(X_n=x_n,T=t)}{P(T=t)}$$
$$=\frac{P(X_n=x_n)P(T-X_n=t-x_n)}{\binom{N}{t}p^tq^{N-t}} \tag{9-126}$$

其中例 9-20 中的方法被用来简化分子项。则

$$P(X_n=x_n|T=t)=\frac{p^{x_n}q^{1-x_n}\binom{N-1}{t-x_n}p^{t-x_n}q^{N-1-t+x_n}}{\binom{N}{t}p^tq^{N-t}}=\frac{\binom{N-1}{t-x_n}}{\binom{N}{t}} \tag{9-127}$$

$$\mathcal{E}[X_n|T=t]=\frac{1}{\binom{N}{t}}\sum_{x_n=0}^{1}x_n\binom{N-1}{t-x_n}=\frac{\binom{N-1}{t-1}}{\binom{N}{t}}=\frac{t}{N} \tag{9-128}$$

它给出了与第一种方法同样的估计器：$\mathcal{E}[X_n|T]=T/N=\overline{X}$。◀

例 9-22 设$\{X_n\}$为 iid 随机变量，具有以下均匀分布：

$$f_X(x)=(1/\theta)I_{[0,\theta)}(x) \tag{9-129}$$

其中由例 9-18($a=0$)，我们知道 $X_{(N)}$ 是 θ 的一个完备充分统计量，它具有以下 pdf：

$$f_{X_{(N)}}(x)=\frac{Nx^{N-1}}{\theta^N}I_{[0,\theta]}(x) \tag{9-130}$$

用第一种方法寻找 $c(\theta)=\theta$ 的 UMVU 估计器，我们计算以下期望值：

$$\mathcal{E}[g(X_{(N)})]=\frac{N}{\theta^N}\int_0^\theta g(x)x^{N-1}\mathrm{d}x \tag{9-131}$$

为了让这个表达式等于 θ，积分必须为 θ^{N+1}/N。选定 $g(x)=x$ 得到：

$$\mathcal{E}[g(X_{(N)})]=\frac{N}{\theta^N}\int_0^\theta x^N\mathrm{d}x=\frac{N\theta^{N+1}}{(N+1)\theta^N}=\frac{N}{N+1}\theta \tag{9-132}$$

则 θ 的 UMVU 估计器为：

$$g(X_{(N)})=\frac{N+1}{N}X_{(N)} \tag{9-133}$$ ◀

在本章的后续部分，我们考虑通过几种其他的方法从 N 个 iid 样本 $\{X_1,\cdots,X_N\}$ 集估计参数 θ。这些估计器也是基于充分统计量，但它们是通过引入一个风险函数推导得到的，这个风险函数可用来产生好的估计器，但可能不同于 UMVU 估计器。它不需要首先找到一个充分统计量，因为对风险函数的优化可直接得到估计器。但是，一个充分统计量通常是最后的结果。我们从经典的贝叶斯估计开始。

9.10 贝叶斯估计

由于θ的估计器 T 是样本$\boldsymbol{X}$ 的一个函数，它是一个随机变量，特征由其分布决定，分布则依赖于 iid 样本的 pdf。除了考虑对θ的充分性，我们对无偏性和有效性也感兴趣，这部分的内容会在本章稍后讨论。我们引入一些用来描述估计器质量的定义。

定义(损失函数) 损失函数 $C(t,\theta)$ 将 $T=t$ 和 θ 映射为一个实数。

有许多的损失函数可用，比如以下的例子：

平方损失：
$$C(t,\theta)=(\theta-t)^2 \tag{9-134}$$

加权平方损失：
$$C(t,\theta)=w(\theta)(\theta-t)^2 \tag{9-135}$$

绝对值损失：
$$C(t,\theta)=|\theta-t| \tag{9-136}$$

门限：
$$C(t,\theta)=\begin{cases}0, & |\theta-t|\leqslant c\\ 1, & |\theta-t|>c\end{cases}=u(|\theta-t|-c) \tag{9-137}$$

Huber：
$$C(t,\theta)=\begin{cases}(\theta-t)^2, & |\theta-t|\leqslant c\\ 2c|\theta-t|-c^2, & |\theta-t|>c\end{cases} \tag{9-138}$$

其中 $w(\theta)$ 是一个加权函数，只与 θ 相关，$c>0$ 是一个门限。图 9-4 中说明了这些损失函数。$C(t,\theta)$ 是衡量 t 偏离真实的 θ 的程度的一个量度。Huber 损失函数在门限 c 处由平方损失变为绝对值损失。一般希望 $C(\cdot)$ 是个严格凸函数(见附录 B)，这样估计器是唯一的。平方损失函数对与 θ 的偏离的评价要比其他函数更严；它是实际中用的最多的损失函数，因为通常用其他的损失函数分析和推导要困难的多。

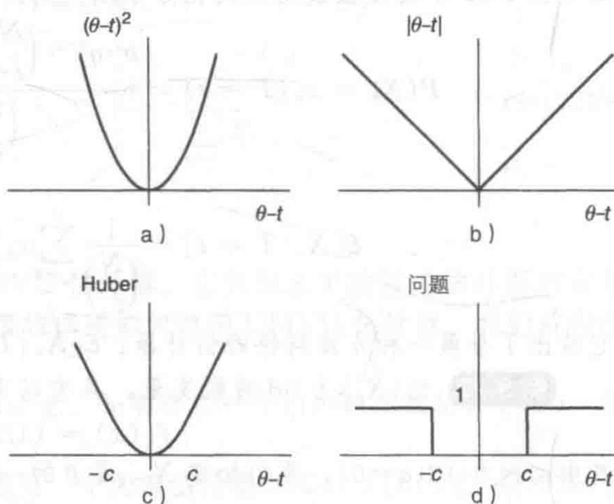


图 9-4　损失函数。a) 平方误差；b) 绝对值误差；c) Huber；d) 门限

定义(风险函数)　随机变量 Θ 的**风险函数**是损失函数的期望值：

$$R\triangleq\mathcal{E}[C(T,\Theta)] \tag{9-139}$$

其中期望值是基于 Θ 和包含 T 的随机变量 $\boldsymbol{X}$ 的联合分布。它也被称为**代价函数**和**目标函数**。

在贝叶斯估计中，通常假设参数为一个随机变量，因此它具有分布 $f_\Theta(\theta)$。目标是找到样本的统计量 T 使得风险函数最小。将此期望展开，R 可被写为以下形式：

$$\begin{aligned}\mathcal{E}[C(T,\Theta)]&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}C(t,\Theta)f_{\Theta,\boldsymbol{X}}(\theta,\boldsymbol{x})\,\mathrm{d}\theta\mathrm{d}\boldsymbol{x}\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}C(t,\Theta)f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)f_\Theta(\theta)\,\mathrm{d}\theta\mathrm{d}\boldsymbol{x}\end{aligned} \tag{9-140}$$

其中第二个表达式是以 Θ 为条件的。注意到上式中 T 是唯一的变量；条件 pdf $f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)$ 和边缘 pdf $f_\Theta(\theta)$ 由特定的问题决定。接下来，我们将介绍一些估计中用在两种重要分布中的专用术语。

定义(先验和后验分布)　未知参数的边缘 pdf $f_\Theta(\theta)$ 叫做**先验分布**。也被称为**优先分布**。**后验分布**是条件 pdf $f_{\Theta|X_1,\cdots,X_N}(\theta|x_1,\cdots,x_N)$。它可以被看做给定样本 $\{X_1,\cdots,X_N\}$ 时关于未知随机参数 Θ 的信息的一个"测度"。

例 9-23　设一个高斯随机变量中均值 μ 为未知参数，σ 已知。由于 N 个样本是 iid 的，条件 pdf 为：

$$\begin{aligned}f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)&=\prod_{n=1}^{N}f_{X_n|\Theta}(x_n|\theta)\\&=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left(-\sum_{n=1}^{N}(x_n-\theta)^2/2\sigma^2\right)\end{aligned} \tag{9-141}$$

这是 N 个已知均值为 θ(因为条件设置)和已知 σ(假设已知)的独立高斯随机变量的联合 pdf。为了继续分析，需要知道 Θ 的先验分布；假设它是均匀的：

$$f_\Theta(\theta)=(1/2)I_{[-1,1]}(\theta) \tag{9-142}$$

则式(9-140)变为：

$$\mathcal{E}[C(T,\Theta)]=\frac{1}{2(2\pi\sigma^2)^{N/2}}\int_{-1}^{1}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}C(t,\theta)$$

$$\cdot \exp\left(-\sum_{n=1}^{N}(x_n-\theta)^2/2\sigma^2\right)\mathrm{d}\theta\mathrm{d}\boldsymbol{x} \tag{9-143}$$

为了继续问题的分析，需要选择一个特定的损失函数 $C(t,\theta)$。目的是找到 T 使得 $\mathcal{E}[C(T,\Theta)]$ 最小，它必须是样本 $\{X_1,\cdots,X_N\}$ 的一个函数。◀

我们稍后再回来讨论这个例题，先分析一些特定的估计技术。

9.11　均方误差估计

考虑两个随机变量 X 和 Y，它们是相关的，联合 pdf 为 $f_{X,Y}(x,y)$。假设 X 是可测量的，而 Y 是无法直接得到的。一个简单的例子是：

$$X=Y+V \tag{9-144}$$

其中 V 是不可测的，并且与 Y 独立。这个模型对许多问题都非常有用，我们需要对 Y 进行估计，但是附加的噪声 V 破坏了样本。由第4章，我们知道当 Y 和 V 独立时(实际问题中通常是这种情况)，X 的 pdf 可由以下卷积关系得到：

$$f_X(x)=\int_{-\infty}^{\infty}f_Y(x-v)f_V(v)\mathrm{d}v \tag{9-145}$$

为了由 X 估计 Y，需要知道联合 pdf $f_{X,Y}(x,y)$。在9.10节的记号中，未知"参数"是 $\Theta=Y$，并且只有一个样本 X，因此 $N=1$。对以上的简单模型，联合 pdf 为

$$f_{X,Y}(x,y)=f_{X|Y}(x|y)f_Y(y) \tag{9-146}$$

运用第4章中的技术：

$$P(X\leqslant x|Y=y)=P(Y+V\leqslant x|Y=y)=P(V\leqslant x-y|Y=y)=P(V\leqslant x-y) \tag{9-147}$$

在最后一个表达式中我们运用了 Y 和 V 独立这样一个事实。则

$$f_{X|Y}(x|y)=\frac{\partial}{\partial x}F_V(x-y)=f_V(x-y) \tag{9-148}$$

$$f_{X,Y}(x,y)=f_V(x-y)f_Y(y) \tag{9-149}$$

回到一般性的从 X 估计 Y 的问题中，最小均方误差(MSE)估计的目的是找到可测随机变量 X 的函数 $h(X)=T$ 使以下的风险函数最小：

$$R\triangleq\mathcal{E}[(Y-h(X))^2] \tag{9-150}$$

为了标记方便，我们用 ξ_{MSE} 来代表 R，工程中常用这个符号来表示 MSE。通常，$h(X)$是一个非线性函数。以线性函数为条件的情况稍后考虑，在 MSE 估计问题中，约束 $h(X)$为线性函数时，我们称其为线性 MSE 估计。由第5章的结果，上式的期望值可以写为：

$$\xi_{\mathrm{MSE}}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-h(x)]^2f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \tag{9-151}$$

它可用条件概率表示如下：

$$\begin{aligned}\xi_{\mathrm{MSE}}&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-h(x)]^2f_{Y|X}(y|x)f_X(x)\mathrm{d}y\mathrm{d}x\\&=\int_{-\infty}^{\infty}f_X(x)\int_{-\infty}^{\infty}[y-h(x)]^2f_{Y|X}(y|x)\mathrm{d}y\mathrm{d}x\end{aligned} \tag{9-152}$$

观察到由于 $f_X(x)\geqslant 0$，使 ξ_{MSE} 最小化等价于使内积分对每个 x 最小。因此，估计问题简化为：

$$h_{\mathrm{MSE}}(x)=\arg\min_{h(\cdot)}\int_{-\infty}^{\infty}[y-h(x)]^2f_{Y|X}(y|x)\mathrm{d}y \tag{9-153}$$

对每个 x 值成立。将积分项对 $h(x)$微分，并令其为0得到：

$$-2\int_{-\infty}^{\infty}yf_{Y|X}(y|x)\mathrm{d}y+2h_{\mathrm{MSE}}(x)\int_{-\infty}^{\infty}f_{Y|X}(y|x)\mathrm{d}y=0 \tag{9-154}$$

由于第二个积分项为 1(因为它是关于 y 的一个有效的 pdf)，我们得到：

$$h_{\mathrm{MSE}}(x)=\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\mathrm{d}y=\mathcal{E}[Y|X=x] \tag{9-155}$$

MSE 估计器是在给定 $X=x$ 时 Y 的条件均值。由于这个表达式对所有 X 的值都成立，因为 X 是随机的，MSE 估计器可被写为

$$h_{\mathrm{MSE}}(X)=\mathcal{E}[Y|X] \tag{9-156}$$

它本身也是一个随机变量，因为它是 X 的一个函数。

MSE 估计问题具有以下性质：

- 最小。式(9-154)对 $h_{\mathrm{MS}}(x)$ 微分得到：

$$\frac{\partial^2 \xi_{\mathrm{MSE}}}{\partial h^2(x)}=2 \tag{9-157}$$

  这意味着 ξ_{MSE} 是凸的，因此 $h_{\mathrm{MSE}}(x)$ 对应于最小的 ξ_{MSE}(不是最大的)。
- 独立性。如果 X 和 Y 独立，则 $\mathcal{E}[Y|X]=\mathcal{E}[Y]$。由于 X 没有提供任何关于 Y 的信息，MSE 估计器简化为 Y 的均值。
- 最小 MSE。如果将 $h_{\mathrm{MSE}}(x)$ 代入式(9-151)，相应的最小 MSE(MMSE)为：

$$\xi_{\mathrm{MSE,min}}=\int_{-\infty}^{\infty} f_X(x)\int_{-\infty}^{\infty}(y-\mathcal{E}[Y|x])^2 f_{Y|X}(y|x)\mathrm{d}y\mathrm{d}x \tag{9-158}$$

$$=\int_{-\infty}^{\infty}\mathrm{var}(Y|x)f_X(x)\mathrm{d}x \tag{9-159}$$

  它是 Y 的条件方差对所有 X 的值取平均。
- 贝叶斯改进。在许多问题中，很难直接计算 $f_{Y|X}(y|x)$。对这种情况，用贝叶斯公式将条件 pdf 重写如下会有利于问题分析：

$$h_{\mathrm{MSE}}(x)=\frac{\int_{-\infty}^{\infty} y f_{X|Y}(x|y)f_Y(y)\mathrm{d}y}{f_X(x)}=\frac{\int_{-\infty}^{\infty} y f_{X|Y}(x|y)f_Y(y)\mathrm{d}y}{\int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)\mathrm{d}y} \tag{9-160}$$

  这种形式计算起来容易些，因为条件是基于不可测的随机变量。在以前的加性噪声模型中，例 9-24 已证明了可以直接找到 $f_{X|Y}(x|y)$的值。

例 9-24 设 $X=Y+V$，假设有以下 pdf：

$$f_Y(x)=(1/2)[\delta(y-1)+\delta(y+1)] \tag{9-161}$$

$$f_V(v)=\frac{1}{\sqrt{2\pi\sigma_V^2}}\exp(-v^2/2\sigma_V^2) \tag{9-162}$$

则 X 是一个对称伯努利随机变量(输出为±1)和一个高斯随机变量(参数为 $\{\mu=0,\sigma_V\}$)的和。由式(9-148)的结果，我们知道

$$f_{X|Y}(x|y)=f_V(x-y)=\frac{1}{\sqrt{2\pi\sigma_V^2}}\exp(-(x-y)^2/2\sigma_V^2) \tag{9-163}$$

$y=\pm1$ 时这个条件 pdf 以及 $f_X(x)$如图 9-5 所示。图 9-6a 则是对应一定范围内 y 值的条件 pdf 的三维图，当考察联合概率密度 $f_{X,Y}(x,y)$ 时非常有用。式(9-160)中的条件均值估计器的分子，可以写做：

$$f_{X|Y}(x|y)f_Y(y)=\frac{1}{2\sqrt{2\pi\sigma_V^2}}[\exp(-(x-1)^2/2\sigma_V^2)\delta(y-1)$$
$$+\exp(-(x+1)^2/2\sigma_V^2)\delta(y+1)] \tag{9-164}$$

其中用到了狄拉克 δ 函数的采样特性。这个表达式是联合 pdf $f_{X,Y}(x,y)$，如图 9-6b 所示。注意到理论上由于 x 是变化的，图中应该有一个狄拉克 δ 函数连续区，这个区域是由式(9-164)中高斯分量的值给定的。

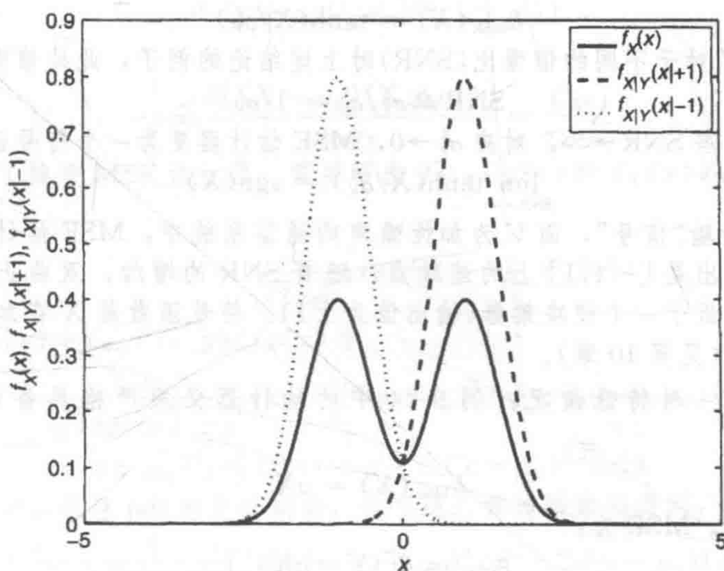


图 9-5　例 9-24 中的条件和边缘 pdf

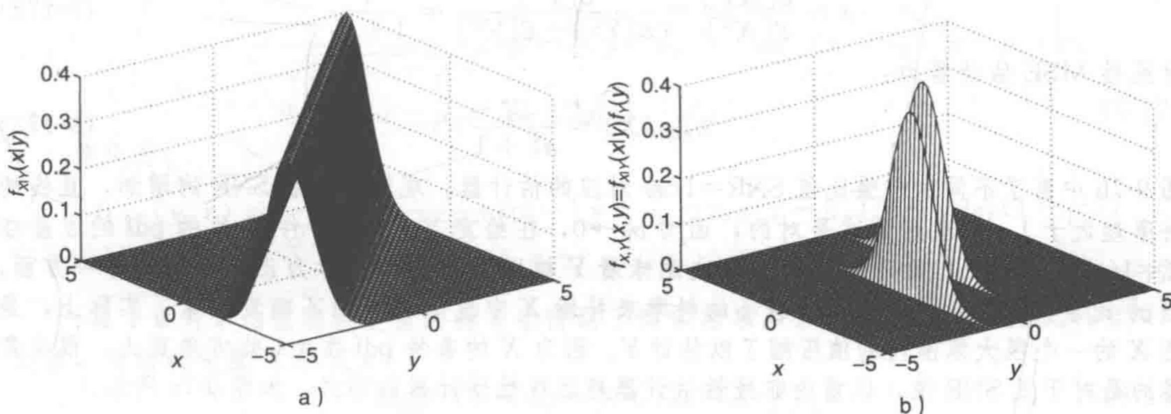


图 9-6　例 9-24 的 pdf。a) 对应一定范围内 y 值的 $f_{X|Y}(x|y)$；b) $f_{X,Y}(x,y)=f_{X|Y}(x|y)f_Y(y)$ 显示了(a)图中只有在 $y=\pm1$ 处由于狄拉克 δ 函数而被保留的两个“切片”

对(9-160)的分子积分得到：

$$\begin{aligned}\int_{-\infty}^{\infty} y f_{X|Y}(x|y) f_Y(y)\mathrm{d}y &= \frac{1}{2\sqrt{2\pi\sigma_V^2}}\left[\exp(-(x-1)^2/2\sigma_V^2)-\exp(-(x+1)^2/2\sigma_V^2)\right]\\ &= \frac{1}{2\sqrt{2\pi\sigma_V^2}}\exp(-(x^2+1)/2\sigma_V^2)\left[\exp(x/\sigma_V^2)-\exp(x/\sigma_V^2)\right]\\ &= \frac{1}{\sqrt{2\pi\sigma_V^2}}\exp(-(x^2+1)/2\sigma_V^2)\sinh(x/\sigma_V^2)\end{aligned} \tag{9-165}$$

其中用到了狄拉克 δ 函数的卷积特性。同样，式(9-160)的分母为：

$$\begin{aligned}\int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\mathrm{d}y &= \frac{1}{2\sqrt{2\pi\sigma_V^2}}\left[\exp(-(x-1)^2/2\sigma_V^2)+\exp(-(x+1)^2/2\sigma_V^2)\right]\\ &= \frac{1}{2\sqrt{2\pi\sigma_V^2}}\exp(-(x^2+1)/2\sigma_V^2)\left[\exp(x/\sigma_V^2)+\exp(-x/\sigma_V^2)\right]\\ &= \frac{1}{\sqrt{2\pi\sigma_V^2}}\exp(-(x^2+1)/2\sigma_V^2)\cosh(x/\sigma_V^2)\end{aligned} \tag{9-166}$$

则 MSE 估计器简化为：

$$h_{\mathrm{MSE}}(X)=\tanh(X/\sigma_V^2) \tag{9-167}$$

图 9-7a 提供了对于不同的信噪比(SNR)时上述结论的例子，此处信噪比的定义为：

$$\mathrm{SNR}\triangleq\sigma_Y^2/\sigma_V^2=1/\sigma_V^2 \tag{9-168}$$

观察极限值，当 SNR→∞，对应 $\sigma_V^2\to 0$，MSE 估计器变为一个符号函数：

$$\lim_{\mathrm{SNR}\to\infty}\tanh(X/\sigma_V^2)=\mathrm{sgn}(X) \tag{9-169}$$

在一个 Y 为传输“信号”，而 V 为加性噪声的通信系统中，MSE 估计器可被看做一个软决策器，它的输出是[−1,1]上的连续点。随着 SNR 的增加，双曲正切显得更象符号函数，在极限处趋近于一个硬决策器(输出值为±1)。符号函数是 X 在加性高斯白噪声条件下的最优检测器(见第 10 章)。◀

例 9-25 考虑一种特殊情况：例 9-24 中的估计器必须严格具备以下简单的线性形式：

$$h_{\mathrm{MSE}}(X)=aX \tag{9-170}$$

其中 a 是待估参量。MSE 为：

$$\xi_{\mathrm{MSE}}=\mathcal{E}[(Y-aX)^2] \tag{9-171}$$

为使其最小，当

$$a=\frac{\mathcal{E}[XY]}{\mathcal{E}[X^2]}=\frac{\mathcal{E}[Y^2]}{\mathcal{E}[Y^2]+\mathcal{E}[V^2]}=\frac{1}{1+\sigma_V^2} \tag{9-172}$$

时线性 MSE 估计器为：

$$h_{\mathrm{MSE}}(X)=\frac{X}{\sigma_V^2+1} \tag{9-173}$$

图 9-7b 中画了不同的信噪比值 SNR$=1/\sigma_V^2$ 对应的估计器。观察到随着 SNR 的增加，直线的斜率趋近于 1。直观感觉这是对的，因为 $\sigma_V^2\to 0$，在给定 $Y=\pm 1$ 条件下 X 的 pdf 的方差在 $X=1$(或者 $X=-1$)附近越来越小，这意味着 Y 被以更大概率估计为正确的值。另一方面，当 σ_V^2 比较大时，估计器用一个较小的斜率来补偿 X 中这个较大的不确定因素；实际上，是将 X 的一个很大范围内的值压缩了以估计 Y。因为 X 的条件 pdf 在±1 处方差更大。很有意思的是对于低 SNR 值，以前的非线性估计器趋近线性估计器的形式，如图 9-7a 所示。

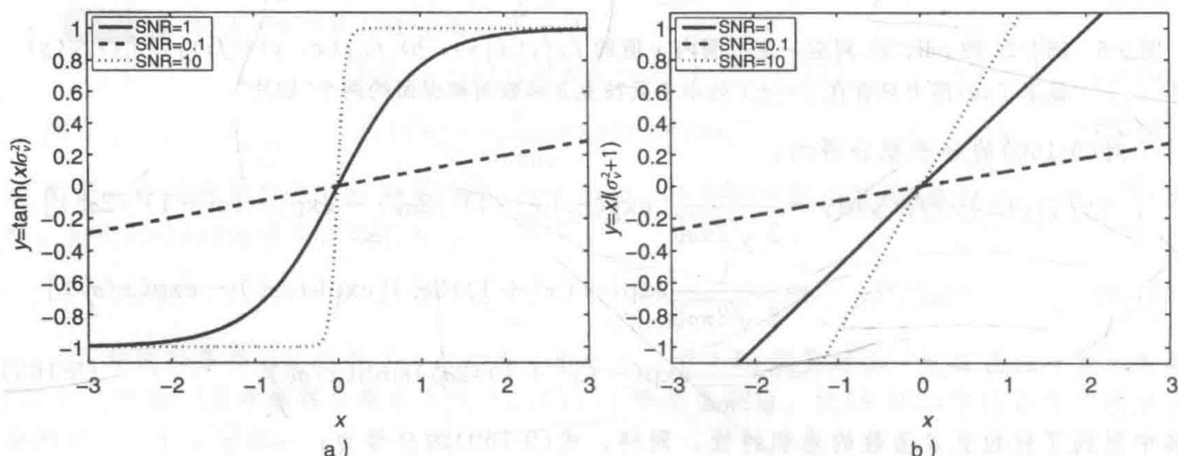


图 9-7　对应不同 SNR 的 MSE 估计器的例子。a) 例 9-24 中非线性估计器；b) 例 9-25 中的线性估计器 ◀

在例 9-25 中，估计器不是条件均值估计器，因为用到了对 $h(X)$的函数形式的约束。在第 11 章，我们考虑扩展这种线性估计器成为作用于随机过程的线性滤波器。

例 9-26 设 X 和 Y 为联合高斯的，联合 pdf 为：

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp(-[(x-\mu_X)^2/\sigma_X^2-2\rho(x-\mu_X)(y-\mu_Y)/\sigma_X\sigma_Y$$

$$+(y-\mu_Y)^2/\sigma_Y^2]/2(1-\rho^2)) \tag{9-174}$$

其中

$$\rho \triangleq \frac{\mathcal{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y} = \frac{C_{XY}}{\sigma_X\sigma_Y} \tag{9-175}$$

为相关系数。为了推导 MSE 估计器，需要写出 $f_{Y|X}(y|x)$和 $f_X(x)$的表达式。很快得到 X 的边缘 pdf 为：

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}}\exp(-(x-\mu_X)^2/2\sigma_X^2) \tag{9-176}$$

条件 pdf $f_{Y|X}(y|x)=f_{X,Y}(x, y)/f_X(x)$为

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\exp(-[(y-\mu_Y)^2/\sigma_Y^2-2\rho(y-\mu_Y)(x-\mu_X)/\sigma_X\sigma_Y$$
$$+(x-\mu_X)^2/\sigma_X^2]/2(1-\rho^2)+(x-\mu_X)^2/2\sigma_X^2) \tag{9-177}$$

正如所期望的，条件 pdf 也是高斯的，这可通过重排指数项得到：

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\exp(-[y-\mu_Y)^2/\sigma_Y^2-2\rho(y-\mu_Y)(x-\mu_X)/\sigma_X\sigma_Y$$
$$+(x-\mu_X)^2/\sigma_X^2-(1-\rho^2)(x-\mu_X)^2/\sigma_X^2]/2(1-\rho^2))$$
$$= \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\exp(-[(y-\mu_Y)^2-2\rho\sigma_Y(x-\mu_X)(y-\mu_Y)/\sigma_X$$
$$+\rho^2\sigma_Y^2(x-\mu_X)^2/\sigma_X^2]/2\sigma_Y^2(1-\rho^2)) \tag{9-178}$$

简化为：

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\exp(-[y-\mu_Y-\sigma_Y\rho(x-\mu_X)/\sigma_X]^2/2\sigma_Y^2(1-\rho^2)) \tag{9-179}$$

（第 4 章对零均值随机变量的推导也得到了类似结果。）条件均值估计器是式(9-179)的均值，用随机变量 X 取代 x：

$$h_{\text{MSB}}(X) = \mu_Y + \rho(\sigma_Y/\sigma_X)(X-\mu_X) \tag{9-180}$$

条件方差和 MMSE 是式(9-179)的方差：

$$\xi_{\text{MSE,min}} = \sigma_Y^2(1-\rho^2) \tag{9-181}$$

因此，当随机变量是联合高斯的，MSE 估计器是一个 X 的仿射函数。考虑以下特殊情况：(i)如果两个随机变量都是零均值，则 $h_{\text{MSE}}(X)=\rho(\sigma_Y/\sigma_X)X$ 恰好是线性的；(ii)如果随机变量不相关，$\rho=0$，则 $h_{\text{MS}}(X)=\mu_Y$ 与 X 独立，MMSE 为 σ_Y^2。◀

9.12 平均绝对误差估计

考虑相关随机变量 X 和 Y，联合 pdf 为 $f_{X,Y}(x,y)$，X 是可测的。平均绝对误差(MAE)风险函数为：

$$\xi_{\text{MAE}} = \mathcal{E}[|Y-h(X)|] \tag{9-182}$$

要将其最小化，我们只需要再次检测以下表达式的内部积分项：

$$\xi_{\text{MAE}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|y-h(x)|f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$
$$= \int_{-\infty}^{\infty}f_X(x)\int_{-\infty}^{\infty}|y-h(x)|f_{Y|X}(y|x)\mathrm{d}x\mathrm{d}y \tag{9-183}$$

绝对值可通过将积分项分为两个积分式子的和来处理，这样估计器可以表示成以下式子：

$$h_{\text{MAF}}(x) = \arg\min_{h(x)}\int_{-\infty}^{\infty}|y-h(x)|f_{Y|X}(y|x)\mathrm{d}y$$

$$=\arg\min_{h(x)}\left[\int_{-\infty}^{h(x)}(h(x)-y)f_{Y|X}(y|x)\mathrm{d}y\right.$$
$$\left.+\int_{h(x)}^{\infty}(y-h(x))f_{Y|X}(y|x)\mathrm{d}y\right] \tag{9-184}$$

要使其最小，只需要将积分项对 $h(x)$ 微分并采用莱布尼茨积分法则(见附录 E)：

$$\frac{\partial}{\partial h(x)}\int_{-\infty}^{h(x)}(h(x)-y)f_{Y|X}(y|x)\mathrm{d}y=\int_{-\infty}^{h(x)}f_{Y|X}(y|x)\mathrm{d}y \tag{9-185}$$

$$\frac{\partial}{\partial h(x)}\int_{h(x)}^{\infty}(y-h(x))f_{Y|X}(y|x)\mathrm{d}y=\int_{h(x)}^{\infty}-f_{Y|X}(y|x)\mathrm{d}y \tag{9-186}$$

其中，$f_{Y|X}(y|x)$不依赖于 $h(x)$。设等号右边项的和为零，我们看到使 MAE 最小的函数是条件中位数：

$$\int_{-\infty}^{h_{\mathrm{MAE}}(x)}f_{Y|X}(y|x)\mathrm{d}y=\int_{h_{\mathrm{MAE}}(x)}^{\infty}f_{Y|X}(y|x)\mathrm{d}y \tag{9-187}$$

这意味着

$$P(Y\leqslant h_{\mathrm{MAE}}(x)\mid X=x)=P(Y\geqslant h_{\mathrm{MAE}}(x)\mid X=x)=1/2 \tag{9-188}$$

对 MSE 估计器也是同样的情况，但并不总是很容易就找到 $f_{Y|X}(y|x)$；用如下所示贝叶斯公式可将 MAE 最优条件重写为：

$$\int_{-\infty}^{h_{\mathrm{MAE}}(x)}f_{X|Y}(x|y)f_Y(y)\mathrm{d}y=\int_{h_{\mathrm{MAE}}(x)}^{\infty}f_{X|Y}(x|y)f_Y(y)\mathrm{d}y \tag{9-189}$$

其中两边的 $f_X(x)$都被消去，因为只对 y 积分。这个结果也可通过联合 pdf 对 y 积分得到：

$$\int_{-\infty}^{h_{\mathrm{MAE}}(x)}f_{X,Y}(x,y)\mathrm{d}y=\int_{h_{\mathrm{MAE}}(x)}^{\infty}f_{X,Y}(x,y)\mathrm{d}y \tag{9-190}$$

例 9-27　对式(9-144)中的加性噪声模型，如例 9-24 中 Y 为对称伯努利以及 V 是高斯的情况，将(9-164)代入(9-189)得到：

$$\int_{-\infty}^{h(x)}[\exp(-(x-1)^2/2\sigma_V^2)\delta(y-1)+\exp(-(x+1)^2/2\sigma_V^2)\delta(y+1)]\mathrm{d}y$$
$$=\int_{h(x)}^{\infty}[\exp(-(x-1)^2/2\sigma_V^2)\delta(y-1)+\exp(-(x+1)^2/2\sigma_V^2)\delta(y+1)]\mathrm{d}y \tag{9-191}$$

其中消去了公共项。被积函数相同，包括两个 $y=\pm1$ 处的狄拉克 δ 函数。很显然，只有当 $h(x)\in[-1,1]$ 时两个积分项都是非零的，这样等式左边包括了 $y=-1$ 的 δ 函数，等式右边包括了 $y=1$ 的 δ 函数。这样，由式(9-191)得到：

$$\exp(-(x+1)^2/2\sigma_V^2)=\exp(-(x-1)^2/2\sigma_V^2) \tag{9-192}$$

则式(9-189)不成立，除非 $x=0$(它发生的概率几乎为 0)。选择离散随机变量中位数作为一个整数输出(见第 5 章)，则式(5-187)和(9-191)说明：

$$h_{\mathrm{MAE}}(X)=\mathrm{sgn}(X) \tag{9-193}$$

这个结果在例 9-24 中当 SNR→∞时也得到过，这是稍后将讨论的最大似然估计器。这个例子将在习题 9-15 中进一步讨论。◀

9.13　正交条件

接下来，我们证明条件均值估计器的一个重要特性。正交性在第 5 章中讨论过，并且条件期望的各种特性都已描述过。

定理 9-8　条件均值估计器 $h_{\mathrm{MAE}}(X)$对所有函数 $g(X)$有以下正交特性：

$$\mathcal{E}[(Y-h_{\mathrm{MSE}}(X))g(X)]=0 \tag{9-194}$$

证明：嵌套期望中以 X 为条件得到：

$$\mathcal{E}[(Y-h_{\mathrm{MSE}}(X))g(X)]=\mathcal{E}[\mathcal{E}[(Y-h_{\mathrm{MSE}}(X))g(X)\mid X]]$$

$$=\mathcal{E}[\mathcal{E}[(Y-h_{\text{MSE}}(X))\,|\,X]g(X)]$$
$$=\mathcal{E}[(\mathcal{E}[Y\,|\,X]-h_{\text{MSE}}(X))g(X)] \tag{9-195}$$

在第二行，条件允许我们将 $g(X)$放到求内部期望运算的外面。最后一行我们利用了 $\mathcal{E}[h_{\text{MSE}}(X)\,|\,X]=h_{\text{MSE}}(X)$这个关系式，因为估计器是 X 的一个函数。由于$\mathcal{E}[Y\,|\,X]=h_{\text{MSE}}(X)$最后的表达式为零，则证明完毕。

正交条件可用来推导最优估计器，因此它是充分必要的。假设式(9-194)成立，则对所有的 $g(X)$

$$\mathcal{E}[h_{\text{MSE}}(X)g(X)]=\mathcal{E}[Yg(X)] \tag{9-196}$$

将等号右边的式子以 X 为条件进行嵌套求期望得到：

$$\mathcal{E}[h_{\text{MSE}}(X)g(X)]=\mathcal{E}[\mathcal{E}[Yg(X)\,|\,X]]=\mathcal{E}[\mathcal{E}[Y\,|\,X]g(X)] \tag{9-197}$$

由于等式必须对所有函数 $g(X)$都成立，而这只有在 $h_{\text{MSE}}(X)=\mathcal{E}[Y\,|\,X]$时才会发生。这种基于正交条件的方法在第 5 章中曾用来证明出现在期望下的数目相等。

假设 MSE 估计器中，我们只能得到样本的函数 $g(X)$，而不是 X。我们可能希望知道 $\mathcal{E}[Y\,|\,X]$和$\mathcal{E}[Y\,|\,g(X)]$是如何相关的。首先，得到一个以 X 为条件的嵌套期望：

$$\mathcal{E}[Yg(X)]=\mathcal{E}[\mathcal{E}[Yg(X)\,|\,X]]=\mathcal{E}[\mathcal{E}[Y\,|\,X]g(X)] \tag{9-198}$$

其中外部期望运算是基于 X 的 pdf。接下来，换成以 $g(X)$为条件：

$$\mathcal{E}[Yg(X)]=\mathcal{E}[\mathcal{E}[Yg(X)\,|\,g(X)]]=\mathcal{E}[\mathcal{E}[Y\,|\,g(X)]g(X)] \tag{9-199}$$

其中外部期望运算是基于 $g(X)$的 pdf。由第 5 章我们知道对于 $Z=g(X)$：

$$\int_{-\infty}^{\infty}g(x)f_X(x)\mathrm{d}x=\int_{-\infty}^{\infty}zf_Z(z)\mathrm{d}z \tag{9-200}$$

则式(9-198)和式(9-199)的内部期望值一定相等：

$$\mathcal{E}[Y\,|\,X]=\mathcal{E}[Y\,|\,g(X)] \tag{9-201}$$

当以 X 或某些函数 $g(X)$为条件时，条件均值估计器是相同的。这个结果是第 5 章中描述的条件期望的平滑性的一种形式。

我们还可以用第 4 章中的变换方法来证明$\mathcal{E}[Y\,|\,g(X)]$与$\mathcal{E}[Y\,|\,X]$相等。假设 $g(X)$是单调增长的：

$$\begin{aligned}\mathcal{E}[Y\,|\,g(X)]&=\int_{-\infty}^{\infty}yf_{Y\,|\,g(X)}(y\,|\,g(x))\mathrm{d}y\\&=\frac{1}{f_{g(X)}(g(x))}\int_{-\infty}^{\infty}yf_{g(X)\,|\,Y}(g(x)\,|\,y)f_Y(y)\mathrm{d}y\end{aligned} \tag{9-202}$$

其中用到了贝叶斯公式。代入以下表达式：

$$f_{g(X)\,|\,Y}(g(x)\,|\,y)=f_{X\,|\,Y}(g^{-1}(g(x))\,|\,y)\left|\frac{\mathrm{d}x}{\mathrm{d}g(x)}\right|=\frac{1}{g'(x)}f_{X\,|\,Y}(x\,|\,y) \tag{9-203}$$

得到：

$$\mathcal{E}[Y\,|\,g(X)]=\frac{1}{f_{g(X)}(g(x))g'(x)}\int_{-\infty}^{\infty}yf_{X\,|\,Y}(x\,|\,y)f_Y(y)\mathrm{d}y \tag{9-204}$$

用贝叶斯公式重写上式的结果得到：

$$\begin{aligned}\mathcal{E}[Y\,|\,g(X)]&=\frac{1}{f_{g(X)}(g(x))g'(x)}\int_{-\infty}^{\infty}yf_{Y\,|\,X}(y\,|\,x)f_X(x)\mathrm{d}y\\&=\frac{f_X(x)}{f_{g(X)}(g(x))g'(x)}\,\mathcal{E}[Y\,|\,X]\end{aligned} \tag{9-205}$$

$\mathcal{E}[Y\,|\,X]$前的乘积项与式(9-203)中相似，除了没有以下的条件：

$$f_{g(X)}(g(x))=f_X(x)/g'(x)\Rightarrow\frac{f_X(x)}{f_{g(X)}(g(x))g'(x)}=1 \tag{9-206}$$

证明了$\mathcal{E}[Y\,|\,g(X)]=\mathcal{E}[Y\,|\,X]$。这个结果对一般的变换 $g(X)$也成立，但如第 4 章中讨论过，当函数为非单调时需要做更多的工作。

9.14 估计器的性质

在这一节，我们描述估计器的几个重要特性。由于估计器是样本的函数(样本是随机变量)，它也是一个随机变量(或对于多维的情况是一个随机向量)，因此其分布也非常重要。我们注意到分布通常是不固定的，如果在估计器中包括更多的样本，则估计器的分布的各种特性(比如方差)也会发生变化。典型的情况是估计器的分布随着 N 的增长更集中于参数 θ 处。

例 9-28 设 $\{X_n\}$ 为 iid 高斯随机变量，均值 μ 未知，σ 已知。回顾一下样本均值 $\overline{X}$ 是 μ 的最小充分统计量。可以直接证明 $\overline{X}$ 也是服从高斯分布，参数为 $\{\mu,\sigma/\sqrt{N}\}$。图 9-8 显示的是其 pdf 的例子，说明了估计器是怎样随着 N 的增加而变得更准确(pdf 的宽度减小了)。

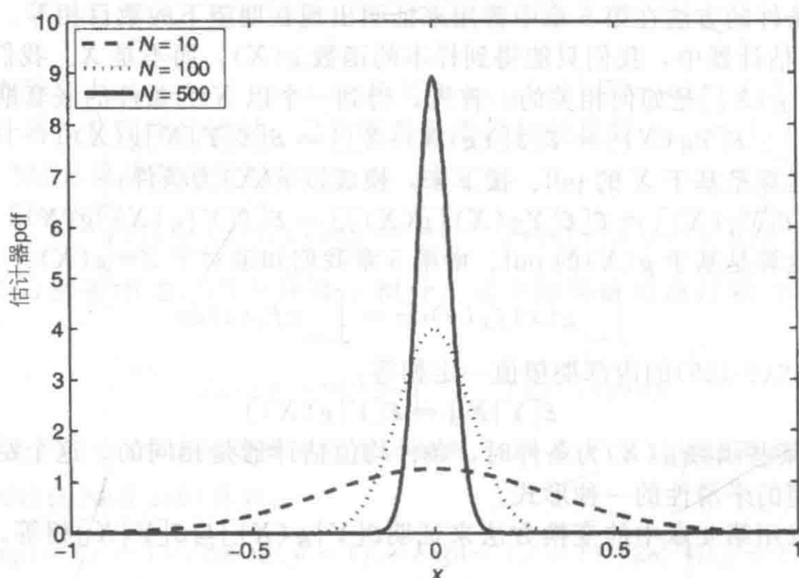


图 9-8 例 9-28 中随着 N 的增大，样本均值 $\overline{X}$ 的 pdf ◀

对于一个估计器的以下特性，假设参数 θ(或参数向量 $\boldsymbol{\theta}$)是非随机的。我们用符号 $\hat{\theta}$ 定义 θ 的估计器来表示它是一个随机变量(代替以前章节中大写的 T)。

9.14.1 无偏性

第一个特性描述的是估计器相对于 θ 的位置。

定义(无偏性) θ 的估计器 $\hat{\theta}$ 是**无偏的**，如果满足 $\mathcal{E}[\hat{\theta}]=\theta$。

无偏性是估计器的重要特性：通常我们希望估计器的均值与参数值相同。但是，不一定必须是这种情况，因为用一个有偏的但是更低方差的估计器可能会更好，只要偏差不是太大。

定义(偏差) 估计器 $\hat{\theta}$ 的偏差 $B(\hat{\theta})$ 为 $B(\hat{\theta})\triangleq\mathcal{E}[\hat{\theta}]-\theta$。

显然，当估计器是无偏的，$B(\hat{\theta})=0$

例 9-29 对于均值为 μ_X 的 iid$\{X_n\}$，$\overline{X}$ 是无偏的：

$$\mathcal{E}[\overline{X}]=\frac{1}{N}\sum_{n=1}^{N}\mathcal{E}[X_n]=\frac{1}{N}\sum_{n=1}^{N}\mu_X=\mu_X \tag{9-207}$$

图 9-8 中例题的 pdf 同样适用这个结论。假设现在我们需要为方差 $\theta=\sigma_X^2$ 设计一个无偏估计器，考虑两种情况：(ⅰ)μ_X 已知；(ⅱ)μ_X 未知。对这两种情况，用到了样本的和，但形式不同：

$$\hat{\theta}=\alpha\sum_{n=1}^{N}(X_n-\mu_X)^2(\mu_X \text{ 已知}) \tag{9-208}$$

$$\hat{\theta}=\alpha\sum_{n=1}^{N}(X_n-\overline{X})^2(\mu_X \text{ 未知}) \tag{9-209}$$

其中在情况(ii)中用到了样本均值，α 是一个系数，用来使每个估计器无偏。对于情况(i)，由于样本是 iid 的，

$$\begin{aligned}\mathcal{E}[\hat{\theta}]&=\alpha\sum_{n=1}^{N}\mathcal{E}[X_n^2-2\mu_X X_n+\mu_X^2]\\&=\alpha N[\sigma_X^2+\mu_X^2-2\mu_X^2+\mu_X^2]=\alpha N\sigma_X^2\end{aligned} \tag{9-210}$$

这表明如果 $\alpha=1/N$，则$\hat{\theta}$是无偏的。情况(ii)计算起来更复杂，因为样本均值 $\overline{X}$：

$$\begin{aligned}\mathcal{E}[\hat{\theta}]&=\alpha\sum_{n=1}^{N}\mathcal{E}[X_n^2-2X_n\overline{X}+\overline{X}^2]\\&=\alpha\Big[N(\sigma_X^2+\mu_X^2)-2\sum_{n=1}^{N}\mathcal{E}[X_n\overline{X}]+N\,\mathcal{E}[\overline{X}^2]\Big]\end{aligned} \tag{9-211}$$

由于样本是 iid 的

$$\mathcal{E}[X_mX_n]=\begin{cases}\mathcal{E}[X_n^2]=\sigma_X^2+\mu_X^2, & m=n\\ \mathcal{E}[X_m]\,\mathcal{E}[X_n]=\mu_X^2, & m\neq n\end{cases} \tag{9-212}$$

因此

$$\sum_{n=1}^{N}\mathcal{E}[X_n\overline{X}]=(1/N)[N(\sigma_X^2+\mu_X^2)+(N^2-N)\mu_X^2] \tag{9-213}$$

同样

$$\mathcal{E}[\overline{X}^2]=\frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\mathcal{E}[X_mX_n]=(1/N^2)[N(\sigma_X^2+\mu_X^2)+(N^2-N)\mu_X^2] \tag{9-214}$$

将式(9-211)中所有项联合得到：

$$\mathcal{E}[\hat{\theta}]=\alpha[N(\sigma_X^2+\mu_X^2)-(1/N)[N(\sigma_X^2+\mu_X^2)+(N^2-N)\mu_X^2]]=\alpha(N-1)\sigma_X^2 \tag{9-215}$$

当 μ_X 未知，如果 $\alpha=1/(N-1)$，则 σ_X^2 的估计器是无偏的。这给出了样本方差 S^2，它在以前章节中讨论过。注意到高斯的假设不是必须的，以前对样本均值的结论是通用的(对 iid 样本)，与特定的分布无关(假设均值存在，方差是有限的)。◀

9.14.2 一致性

接下来，我们想确定随着样本数目 N 的增长，估计器是否会在某种意义下收敛于 θ；这样的估计器被称为具有一致性。

定义(弱一致性) 估计器$\hat{\theta}(N)$是**弱一致**的，如果它依概率收敛：

$$\lim_{N\to\infty}P(|\hat{\theta}-\theta|\geq\varepsilon)=0 \tag{9-216}$$

对每个 $\varepsilon>0$。

尽管一致性意味着$\hat{\theta}$的 pdf 变得越来越集中于 θ，它并不直接专注于估计器的方差是否趋近于零。一致性也可以用几乎必然收敛的方式来描述。

定义(强一致性) 估计器$\hat{\theta}$是**强一致**的，如果它几乎必然收敛：

$$P(\lim_{N\to\infty}\hat{\theta}=\theta)=1 \tag{9-217}$$

如果一个估计器是强一致的，则它也是弱一致的，因为几乎必然收敛意味着依概率收敛，如第7章中所讨论的。

例 9-30 考虑例 9-29 中的第(i)种情况，其中 $\overline{X}$ 用来估计一个已知方差 σ_X^2 的高斯分布的均值 μ_X。由于 $\overline{X}$ 均值 μ_X 和方差 σ_X^2/N

$$P(|\overline{X}-\mu_X|\geqslant\varepsilon)=P(|\overline{Z}|\geqslant\varepsilon)=2\int_{\varepsilon}^{\infty}\frac{1}{\sqrt{2\pi\sigma_X^2/N}}\exp(-z^2/2(\sigma_X^2/N))\mathrm{d}z \tag{9-218}$$

其中 $\overline{Z}\triangleq\overline{X}-\mu_X$ 是零均值高斯的，方差为 σ_X^2/N，我们用到了它的 pdf 是偶函数的特性。根据 $y\triangleq z/(\sigma_X/\sqrt{N})$ 替换变量得到：

$$P(|\overline{X}-\mu_X|\geqslant\varepsilon)=2\int_{\sqrt{N}\varepsilon/\sigma_X}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-y^2/2)\mathrm{d}y=2Q(\sqrt{N}\varepsilon/\sigma_X) \tag{9-219}$$

其中 $Q(\cdot)$是 Q 函数。很显然随着 $N\to\infty$ 对每个 $\varepsilon>0$，以上的概率趋近于零，则 $\overline{X}$ 是弱一致的。当然，这个结论是第 7 章中讨论过的弱大数定理的重述。以上的概率是图 9-8中对应 $\mu_X=0$ 时 pdf 尾部下的面积，从图中可以很清晰的看到对于一个固定的 ε，它随着 N 的增长而减小。 ◀

我们也可以用切比雪夫不等式来证明这个估计器是弱一致的(见附录 F)：

$$P(|\overline{X}-\mu_X|\geqslant\varepsilon)\leqslant\frac{\sigma_X^2}{N\varepsilon^2} \tag{9-220}$$

对于正的 ε，随着 $N\to\infty$式子的右边趋于零。根据强大数定理(见第 7 章)，显然 $\overline{X}$ 也是强一致的，但推导起来更复杂。

例 9-31 对例 9-29 中的第(i)种情况，一个 σ_X^2 的无偏估计器的方差是在假设 μ_X 已知后推导出的。由于

$$\mathcal{E}[(\hat{\theta}-\theta)^2]=\mathcal{E}[\hat{\theta}^2]-\theta^2 \tag{9-221}$$

$\hat{\theta}$是无偏的，只需要计算$\mathcal{E}[\hat{\theta}^2]$。因此

$$\mathcal{E}[\hat{\theta}^2]=\frac{1}{N^2}\mathcal{E}\left[\left(\sum_{n=1}^{N}(X_n-\mu_X)^2\right)^2\right]=\frac{1}{N^2}\mathcal{E}\left[\left(\sum_{n=1}^{N}Y_n^2\right)^2\right] \tag{9-222}$$

其中 $Y_n\triangleq X_n-\mu_X$ 具有零均值和方差 σ_X^2。由于$\{Y_n\}$是 iid 的，期望值简化为：

$$\mathcal{E}\left[\left(\sum_{n=1}^{N}Y_n^2\right)^2\right]=N\,\mathcal{E}[Y_n^4]+(N^2-N)\,\mathcal{E}^2[Y_n^2] \tag{9-223}$$

假设 X_n 是高斯的，参数为 $\{\mu_X,\sigma_X\}$，则 Y_n 也是高斯的，这允许我们用方差表示一个四阶中心距(见第 5 章)：

$$\mathcal{E}[Y_n^4]=3\sigma_X^4 \tag{9-224}$$

将以上所有结论联合，代入 $\theta^2=\sigma_X^4$，得到：

$$\mathcal{E}[(\hat{\theta}-\theta)^2]=(1/N^2)[3N\sigma_X^4+(N^2-N)\sigma_X^4]-\sigma_X^4=2\sigma_X^4/N \tag{9-225}$$

当 $N\to\infty$时，上式趋于零。 ◀

例 9-32 对例 9-31 中高斯样本，我们回到已知 μ_X 时 σ_X^2 的常用估计器

$$\hat{\theta}=\alpha\sum_{n=1}^{N}(X_n-\mu_X)^2 \tag{9-226}$$

由式(9-225)的结论，得到式(9-226)的方差：

$$\mathcal{E}[(\hat{\theta}-\theta)^2]=\alpha^2[2N\sigma_X^4+N^2\sigma_X^4]-\sigma_X^4=[\alpha^2N(N+2)-1]\sigma_X^4 \tag{9-227}$$

为了使方差小于式(9-225)中无偏估计器的方差，我们需要

$$\alpha^2N(N+2)-1<2/N\Rightarrow\alpha<1/N \tag{9-228}$$

但是，为了得到一个有效估计器，还有必要满足：

$$\alpha^2N(N+2)-1>0\Rightarrow\alpha>\frac{1}{\sqrt{N(N+2)}} \tag{9-229}$$

最后的结论使我们无法使用 $\alpha=1/(N+1)$。图 9-9 显示了 α 对应 N 的上下界的约束以及 $\alpha=1/(N+1/2)$的关系。在 N 较小时，我们看到用比例因子比用 $\alpha=1/N$ 更有优势，但

增益适中，并且不能用 σ_X^2 的无偏估计器证明。

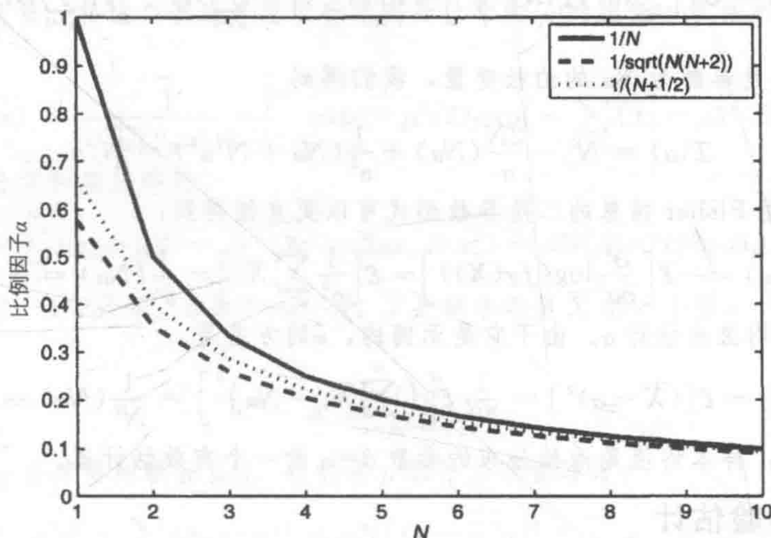


图 9-9　例 9-32 中样本方差的比例因子 α 的值(为了看起来清晰将离散的值用直线连接起来)　◀

9.14.3　有效性

估计器的有效性由其方差决定。

定义(有效性)　一个无偏估计器 $\hat{\theta}$ 是**有效**的，如果它的方差达到克拉美罗下限。

克拉美罗限(CRLB)是 $\boldsymbol{\theta}$ 的协方差矩阵的完整的边界，它由 $\{X_n\}$ 的分布参数和样本的数目 N 确定。在附录 F 中推导得到：

$$\operatorname{cov}[\hat{\boldsymbol{\theta}}] \geqslant \mathcal{I}^{-1}(\boldsymbol{\theta}) \tag{9-230}$$

其中

$$\mathcal{I}(\boldsymbol{\theta}) = \mathcal{E}\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}}\log(f_{\boldsymbol{X}}(\boldsymbol{X})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}}\log(f_{\boldsymbol{X}}(\boldsymbol{X}))\right)^{\mathrm{T}}\right] \tag{9-231}$$

是 Fisher 信息矩阵，$f_{\boldsymbol{X}}(\boldsymbol{X})$ 是样本 $\boldsymbol{X} \triangleq [X_1,\cdots,X_N]^{\mathrm{T}}$ 的联合 pdf，$\boldsymbol{\theta} \triangleq [\theta_1,\cdots,\theta_M]^{\mathrm{T}}$。注意到期望运算中的 pdf 的变量是随机向量 $\boldsymbol{X}$。上面的不等式是在附录 F 中描述的半正定(PSD)意义下定义的。Fisher 信息矩阵的另一种更容易计算的形式是基于二阶导数：

$$\mathcal{I}(\boldsymbol{\theta}) = -\mathcal{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\log(f_{\boldsymbol{X}}(\boldsymbol{X}))\right] \tag{9-232}$$

对数可以是任意底，但通常用 $\ln(f_{\boldsymbol{X}}(\boldsymbol{X}))$。

例 9-33　设 $\{X_1,\cdots,X_N\}$ 是一个泊松随机变量的 iid 样本，未知参数 $\theta=\alpha$，联合 pdf 为：

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\exp(-N\alpha)\alpha^{\sum_{n=1}^{N} x_n}}{\prod_{n=1}^{N} x_n!} \tag{9-233}$$

这个式子的对数为

$$\ln(f_{\boldsymbol{X}}(\boldsymbol{x})) = -N\alpha + \sum_{n=1}^{N} x_n \ln(\alpha) - \ln\left(\prod_{n=1}^{N} x_n!\right) \tag{9-234}$$

微分后为

$$\frac{\partial}{\partial \alpha}\ln(f_{\boldsymbol{X}}(\boldsymbol{x})) = -N + \frac{1}{\alpha}\sum_{n=1}^{N} x_n \tag{9-235}$$

于是(标量)Fisher 信息为

$$\mathcal{I}(\alpha) = \mathcal{E}\left[\left(\frac{\partial}{\partial \alpha}\ln(f_{\boldsymbol{X}}(\boldsymbol{X}))\right)^2\right] = \mathcal{E}\left[N^2 - \frac{2N}{\alpha}\sum_{n=1}^{N} X_n + \frac{1}{\alpha^2}\left(\sum_{n=1}^{N} X_n\right)^2\right] \tag{9-236}$$

由于 $\sum_{n=1}^{N} X_n$ 是参数为 $N\alpha$ 的泊松变量，我们得到

$$\mathcal{I}(\alpha) = N^2 - \frac{2N}{\alpha}(N\alpha) + \frac{1}{\alpha^2}(N\alpha + N^2\alpha^2) = N/\alpha \tag{9-237}$$

CRLBα/N 用 Fisher 信息的二阶导数形式可以更直接得到：

$$\mathcal{I}(\alpha) = -\mathcal{E}\left[\frac{\partial^2}{\partial \alpha^2}\log(f_{\boldsymbol{X}}(\boldsymbol{X}))\right] = \mathcal{E}\left[\frac{1}{\alpha^2}\sum_{n-1}^{N} X_n\right] = \frac{1}{\alpha^2}(N\alpha) = N/\alpha \tag{9-238}$$

考虑用样本均值来估计 α。由于它是无偏的，$\hat{\alpha}$的方差是

$$\text{var}[\hat{\alpha}] = \mathcal{E}[(\overline{X} - \alpha)^2] = \frac{1}{N^2}\mathcal{E}\left[\left(\sum_{n=1}^{N} X_n - N\alpha\right)^2\right] = \frac{1}{N^2}(N\alpha) = \alpha/N \tag{9-239}$$

它达到了 CRLB：样本均值是泊松分布的参数 $\theta=\alpha$ 的一个有效估计器。 ◀

9.15 最大后验估计

假设我们想要从 N 个 iid 样本 $\{X_1,\cdots,X_N\}$ 估计随机参数 Θ，设它具有已知 pdf $f_\Theta(\theta)$。回顾一下在估计问题中，$f_\Theta(\theta)$ 被称为先验分布，它在样本被观察前就描述了 Θ 的特征。后验分布 $f_{\Theta|X_1,\ldots,X_N}(\theta|x_1,\ldots,x_N)$ 则是在收集了 N 个样本后对 Θ 的特征进行描述。用贝叶斯公式，后验分布可由先验分布表示如下：

$$f_{\Theta|\boldsymbol{X}}(\theta|\boldsymbol{x}) = \frac{f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)f_\Theta(\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \tag{9-240}$$

分子是给定 $\Theta=\theta$ 后样本的联合 pdf，而分母是不设置任何条件下的样本的联合 pdf。通常，分子中的两项都假设是已知的。而分母通常是未知的，但是可由分子推导得到，如下所示：

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{\infty} f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)f_\Theta(\theta)\,\mathrm{d}\theta \tag{9-241}$$

最大后验(MAP)估计器 θ_{MAP} 就是使式(9-240)的左边的后验分布最大的值。由于分母与 θ 没有直接的关系，它不影响最大值，可以被忽略。要得到最大值必须联合考虑等式右边分子中两个 pdf。稍后，我们将描述最大似然(ML)估计器，它忽略了 Θ 的先验 pdf，取而代之的是找到 θ_{ML} 使式(9-240)右边(它被称为似然函数)的条件 pdf 最大。在本章的剩余部分，我们将用下标来标示特定类型的估计器(代替之前用到的符号$\hat{\theta}$)。

定义(MAP 估计器) 参数 Θ 的 MAP **估计器**是使后验 pdf 最大的值：

$$\theta_{\text{MAP}} = \arg\max_\theta f_{\Theta|\boldsymbol{X}}(\theta|\boldsymbol{x}) \tag{9-242}$$

由于式(9-240)的分母与 θ 无关，MAP 准则可重写为

$$\theta_{\text{MAP}} = \arg\max_\theta f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)f_\Theta(\theta) = \arg\max_\theta f_{\boldsymbol{X},\Theta}(\boldsymbol{x},\theta) \tag{9-243}$$

通常用后一种形式，因为在给定 Θ 后确定样本的条件 pdf 更容易。当然这个表达式仍然需要我们知道 Θ 的先验 pdf，这在 MAP 估计中是必须的。

在给出一个例子前，我们先对符号做一个说明。在之前涉及充分统计量的章节中，$f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ 被用来表示$\{X_n\}$的联合 pdf，参数为 θ，我们用它来寻找充分统计量。它与条件 pdf $f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)$的记号的相等的，$f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)$以 Θ 为条件，Θ 的所有随机性都被忽略了。但是，对 MAP 估计器，我们需要考虑样本和参数的联合 pdf：$f_{\boldsymbol{X},\Theta}(\boldsymbol{x},\theta)$。

例 9-34 考虑一个高斯随机变量，参数为 $\{\mu,\sigma\}$，已知 σ。假设 μ 是一个标准高斯分布的随机变量。对 N 个 iid 样本 $\{X_1,\cdots X_N\}$，条件 pdf 为：

$$f_{\boldsymbol{X}|\mu}(\boldsymbol{x}|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left(-\sum_{n=1}^{N}(x_n-\mu)^2/2\sigma^2\right) \tag{9-244}$$

（由于大写的 μ 与拉丁字母 M 相同，我们在下标中用小写的 μ，从避免引起混淆，它仍是一个随机变量。）将这个表达式乘以标准高斯分布的 μ 的边缘 pdf 再积分就得到了样本的联合 pdf：

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}(2\pi\sigma^2)^{N/2}}\int_{-\infty}^{\infty}\exp(-\mu^2/2)\exp\Big(-\sum_{n=1}^{N}(x_n-\mu)^2/2\sigma^2\Big)\mathrm{d}\mu \tag{9-245}$$

将指数项合并和展开得到：

$$\mu^2\sigma^2 + \sum_{n=1}^{N}(x_n-\mu)^2 = \sum_{n=1}^{N}(a\mu^2 - 2\mu x_n + x_n^2) = aN[\mu^2 - (2\overline{x}/a)\mu + \overline{x^2}/a] \tag{9-246}$$

其中我们忽略了 $-1/2\sigma^2$ 项，$a \triangleq 1+\sigma^2/N$，$\overline{x}$ 是样本均值 $\overline{X}$ 的一个值，$\overline{x^2}$ 是样本的二阶矩的一个值：

$$\overline{X^2} \triangleq \frac{1}{N}\sum_{n=1}^{N}X_n^2 \tag{9-247}$$

由于是对 μ 合并得到联合 pdf，我们将 μ 的平方项重写为：

$$\mu^2\sigma^2 + \sum_{n=1}^{N}(x_n-\mu)^2 = aN[(\mu-\overline{x}/a)^2 + \overline{x^2}/a - \overline{x}^2/a^2] \tag{9-248}$$

将这个结果代入式(9-245)得到：

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \frac{1}{\sqrt{2\pi}(2\pi\sigma^2)^{N/2}}\exp(-(N/2\sigma^2)(\overline{x^2}-\overline{x}^2/a)) \\ &\quad \cdot \int_{-\infty}^{\infty}\exp(-(aN/2\sigma^2)(\mu_X-\overline{x}/a)^2)\mathrm{d}\mu \\ &= \frac{\sqrt{\sigma^2/aN}}{(2\pi\sigma^2)^{N/2}}\exp(-(N/2\sigma^2)(\overline{x^2}-\overline{x}^2/a)) \end{aligned} \tag{9-249}$$ ◀

积分为 1，因为积分项是均值为$\overline{x^2}/a$，方差为 σ^2/aN 的高斯 pdf。图 9-10a 给出了 $N=2$ 和 $\sigma=1$ 的图。观察到它关于 $x_1=x_2$ 直线对称；这个对称性可以扩展到高维，因为 pdf 是通过 $\overline{x}$ 和$\overline{x^2}$与观测值相关的一个函数，它们不依赖于样本的次序。对 MAP 估计器，这个 pdf 结论并不重要，因为 $f_{\boldsymbol{X}}(\boldsymbol{x})$不是 μ 的函数。以上我们给出推导过程就是为了强调：当 Θ 为随机变量时，联合 pdf 与 $f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)$不同；它是一个$\{x_n\}$的复杂表达式，在这个例子中它甚至不是高斯的。

图 9-10b 显示的是式(9-244)在特定值 $\mu=-1$ 时的条件 pdf $f_{\boldsymbol{X}|\mu}(\boldsymbol{x}|\mu)$。尽管 $f_{\boldsymbol{X}|\mu}(\boldsymbol{x}|\mu)$ 与图 9-10a 中的 pdf 看起来相似，但它实际上是高斯的，因为样本是 iid 的，它关于中心圆对称。由于 μ 是变化的，pdf 的中心沿着 $x_1=x_2$ 直线移动。在图 9-10c 中，我们给出了这两个 pdf 在 $\mu=-1$ 时的比例，它被用于式(9-240)的贝叶斯公式中。由于 μ 是变化的，这个象波浪样的 pdf 的比值沿着 $x_1=x_2$ 直线移动。很有趣的是它的最大值沿着 $x_2+\mu=-(x_1+\mu)$定义的直线出现。

我们再来考虑当 $\theta=\mu$ 变化时式(9-240)的后验分布。等式右边的联合 pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$不随 μ 而改变；对这个例子，它只通过 $\overline{X}$ 和$\overline{X^2}$与样本相关。式(9-240)左边的先验分布是$f_\mu(\mu)$，它是标准高斯 pdf。条件 pdf $f_{\boldsymbol{X}|\mu}(\boldsymbol{x}|\mu)$是 N 个高斯 pdf 的乘积，它们都具有相同的均值 μ。但是由于 μ 是变化的，我们注意到一个有趣的现象：我们可将乘积中的每个边缘 pdf 都看做参数为 μ 的高斯分布，它们均值都不同，由 x_n 给定。（这种样本和参数的二元性对大多数分布并不存在。）由于 μ 是变化的，式(9-240)的右边是 $N+1$ 个高斯 pdf 的乘积：其中一个具有零均值和单位方差，其他的 N 个具有均值$\{x_n\}$和相同的方差 σ^2。等式右边的分子面积不为 1；对 $f_{\boldsymbol{X}|\mu}(\boldsymbol{x}|\mu)$中特定的$\{x_n\}$值，分母 $f_{\boldsymbol{X}}(\boldsymbol{x})$确保等式左边具有单位面积（考虑到 μ）。为了画出条件 pdf 与 μ 的关系，需要$\{x_n\}$的一些值。图 9-11 中显示了 $N=2$ 时三个不同 $\{x_1,x_2\}$ 集对应的先验和后验分布。

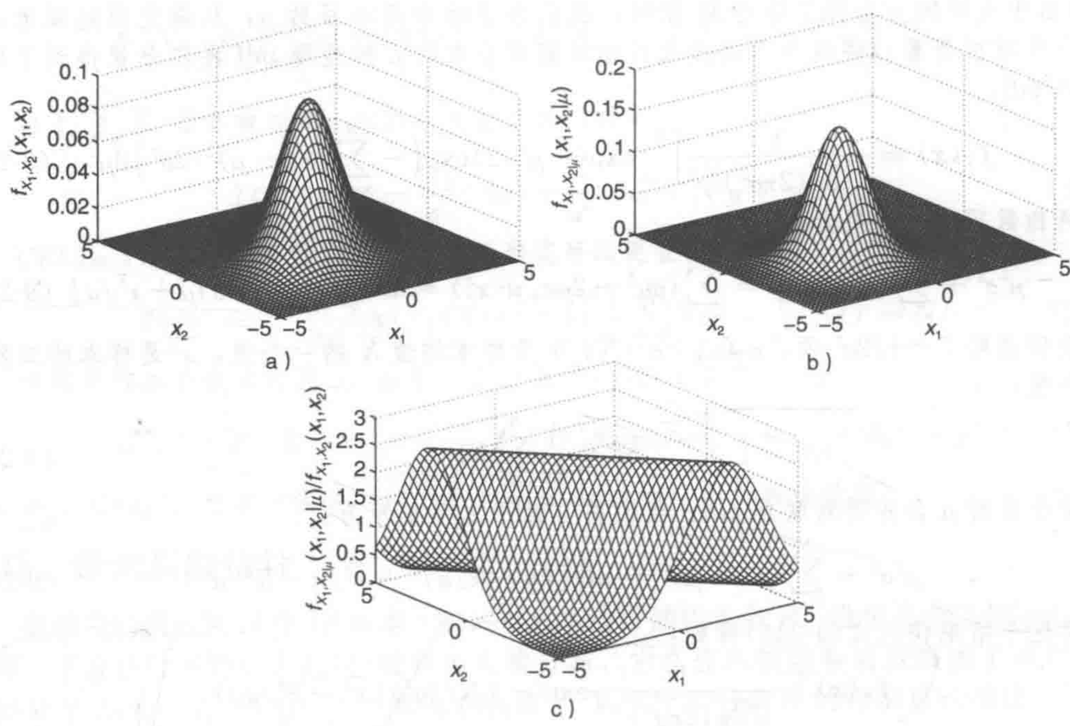


图 9-10　$N=2$ 时例 9-34 的 pdf。a) 联合 pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$；b) $\mu=-1$ 时条件 pdf $f_{\boldsymbol{X}|\mu}(\boldsymbol{x}|\mu)$；c) pdf 的比值

图 9-11　例 9-34 中先验和后验 pdf，后验 pdf 的最大值分别发生在 0，$-5/3$ 和 $-10/3$ 处。a) $x_1=x_2=0$；b) $x_1=0$，$x_2=-5$；c) $x_1=x_2=-5$

当然，后验分布依赖于特定的样本值。尽管由先验分布知道随机变量 μ 的均值为零，非零样本将后验 pdf 向那些样本移动。但是，当 N 变得很大时，我们希望 $\overline{X}$ 会趋近于 μ 的均值，这样 $f_{\mu|\boldsymbol{X}}(\mu|\boldsymbol{x})$ 就会更集中于零附近，方差就会变得很小。这服从 MAP 准则。由于式(9-240)中分母被忽略(由于它与 μ 无关)，MAP 准则变为：

$$\begin{aligned}\mu_{\mathrm{MAP}} &= \arg\max_{\mu}\frac{1}{\sqrt{2\pi}}\exp(-\mu^2/2)\prod_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x_n-\mu)^2/2\sigma^2)\\ &= \arg\max_{\mu}\exp\Big(-\mu^2/2-(1/2\sigma^2)\sum_{n=1}^{N}(x_n-\mu)^2\Big)\end{aligned}\tag{9-250}$$

其中包括 π 的常数项被忽略了，因为它不对取最大值产生影响。通常对指数函数取对数是比较方便的：由于对数是单调的，在最大值求解过程中得到相同的 μ 值。这样

$$\mu_{\mathrm{MAP}} = \arg\max_{\mu}\Big(-\mu^2/2-(1/2\sigma^2)\sum_{n=1}^{N}(x_n-\mu)^2\Big)\tag{9-251}$$

将这个表达式对 μ 微分，将结果等于零，求解得到：

$$\mu_{\mathrm{MAP}} = \frac{1}{\sigma^2+N}\sum_{n=1}^{N}X_n\tag{9-252}$$

其中我们用到了表达式中的样本$\{X_n\}$，而不是输出$\{x_n\}$。图 9-11 中对三张图中用到的特定的输出，后验 pdf 的最大值出现在$(x_1+x_2)/(\sigma^2+N)=0$，$-5/3$ 和$-10/3$ 处。当方差 σ^2 变大时(对固定的 N)，估计器 μ_{MAP}趋近于 0，即先验分布的均值。当 σ^2 变小时(意味着测量值靠拢后验分布的 μ)，μ_{MAP}趋近于样本均值，我们在 9.16 节看到的就是这个例子的 ML 估计器。

例 9-35 考虑式(9-43)中指数族随机变量的表达式。对 N 个 iid 的样本，MAP 估计准则为

$$\begin{aligned}\theta_{\mathrm{MAP}} &= \arg\max_{\theta}\ln\Big(a^N(\theta)\exp\Big(b(\theta)\sum_{n=1}^{N}c(x_n)\Big)\prod_{n=1}^{N}\mathrm{d}(x_n)f_{\Theta}(\theta)\Big)\\ &= \arg\max_{\theta}\Big[N\ln(a(\theta))+b(\theta)\sum_{n=1}^{N}c(x_n)+\ln(f_{\Theta}(\theta))\Big]\end{aligned}\tag{9-253}$$

其中用到了式(9-243)的对数运算，包含$\{\mathrm{d}(x_n)\}$的项被去掉了，因为它不会影响最大值。假设$\{X_n\}$具有指数分布，由表 9-2 知道样本均值对 $\theta=\lambda$ 是充分的。实际上，它是 UMVU 估计器，因为它达到了 CRLB。在上面的表达式中，$a(\theta)=-b(\theta)=\lambda$，$c(x_n)=x_n$。将其对 λ 微分，将结果置零得到 MAP 估计器的表达式为：

$$N/\lambda_{\mathrm{MAP}}-\sum_{n-1}^{N}X_n+\frac{f'_{\Lambda}(\lambda_{\mathrm{MAP}})}{f_{\Lambda}(\lambda_{\mathrm{MAP}})}=0\tag{9-254}$$

其中 $f'_{\Lambda}(\lambda)$是 λ 的一般导数。设先验分布也是指数的，具有已知参数 λ_0：

$$f_{\Lambda}(\lambda)=\lambda_0\exp(-\lambda_0\lambda)I_{[0,\infty)}(\lambda)\tag{9-255}$$

则

$$\frac{f'_{\Lambda}(\lambda_{\mathrm{MAP}})}{f_{\Lambda}(\lambda_{\mathrm{MAP}})}=-\lambda_0\Rightarrow\lambda_{\mathrm{MAP}}=\frac{1}{N}\Big(\sum_{n=1}^{N}X_n+\lambda_0\Big)=\overline{X}+\lambda_0/N\tag{9-256}$$

先验分布的参数导致估计器偏离了样本均值(即 UMVU 估计器)。先验分布随着 N 的增长，影响逐渐减弱，对小的 λ_0 也同样，因此 $f_{\Lambda}(\lambda)$相对是比较平坦的。◀

如果 MAP 准则中不包括先验分布 $f_{\Theta}(\theta)$，我们有 9.16 节中描述的 ML 准则。由式(9-253)可以很明显看出，当 $f_{\Theta}(\theta)$为均匀分布时，它对 MAP 估计器没有影响，这种情况下 MAP 和 ML 准则相同(这通常是真的，而不仅仅对指数族成立)。

9.16 最大似然估计

考虑估计参数 θ，它的先验分布 $f_{\Theta}(\theta)$是未知的，不能用 MAP 估计。对这种情况，我们可以设计一个与 MAP 估计器相似的估计器，它不需要任何关于先验分布的信息。

定义(似然函数) 与 N 个随机变量 $\{X_1,\cdots,X_N\}$ 相关的**似然函数**是条件 pdf：

$$l(\boldsymbol{x};\theta)\triangleq f_{X|\Theta}(\boldsymbol{x}|\theta) \tag{9-257}$$

对数似然函数为：

$$L(\boldsymbol{x};\theta)\triangleq\log(l(\boldsymbol{x};\theta)) \tag{9-258}$$

其中 log(•)通常是自然对数。

似然函数是针对样本 $\{X_1,\cdots,X_N\}$ 的特定输出时未知参数 θ 的函数，我们感兴趣的是使 $l(\boldsymbol{x};\theta)$ 最大的 θ 的值；它是与特定输出对应的"最可能"的值。我们通常会更多选择对数似然函数，因为对数运算通常会简化表达式，这样最大值更容易找到，就如对指数族的情况一样。由于对数运算是单调递增函数，似然函数和对数似然函数最大时，θ_{ML} 相同。

定义(ML 估计器) 参数 θ 的 **ML 估计器** θ_{ML} 使似然函数 $l(\boldsymbol{x};\theta)$ 和对数似然函数 $L(\boldsymbol{x};\theta)$ 最大。

如果似然函数对 θ 的导数存在，ML 估计器可由求解以下式子导出：

$$\frac{\partial}{\partial\theta}L(\boldsymbol{x};\theta)=0 \tag{9-259}$$

对一个指数族，我们可以用式(9-253)的 MAP 的结论，忽略先验分布得到：

$$\frac{N}{a(\theta)}a'(\theta)+b'(\theta)\sum_{n=1}^{N}c(X_n)=0 \tag{9-260}$$

其中上标表示的是对 θ 求一般导数。对例 9-35 中的指数分布，我们得到 $a(\theta)=\theta$，$a'(\theta)=1$，$b'(\theta)=1$，$c(X_n)=X_n$，因此 $\theta_{ML}=\overline{X}$。如果导数定义不明确，则需要用其他的方法来找到使似然函数最大的 θ。这将在后面分析均匀分布的样本时再证明。

例 9-36 我们想估计已知 σ 的高斯随机变量 X 的均值 μ。对 N 个独立的样本，似然函数为：

$$\begin{aligned}l(\boldsymbol{x};\mu)&=\frac{1}{(2\pi\sigma^2)^{N/2}}\prod_{n=1}^{N}\exp(-1/2\sigma^2(x_n-\mu)^2)\\&=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left(-1/2\sigma^2\sum_{n=1}^{N}(x_n-\mu)^2\right)\end{aligned} \tag{9-261}$$

图 9-12 中显示的是 $N=2$ 以及与图 9-11 中用到的相同的 $\{x_1,x_2\}$ 值的似然函数的例子。注意到参数 $\theta=\mu$ 时似然函数不是一个 pdf。为了更容易对图中三个曲线进行比较，对它们都做了归一化，使最大值为 1。观察到每条曲线的峰值出现在两个样本值的平均值处：0，−2.5 和 −5 分别对应 $\{x_1,x_2\}=\{0,0\}$，$\{0,-5\}$ 和 $\{-5,-5\}$。对数似然函数为：

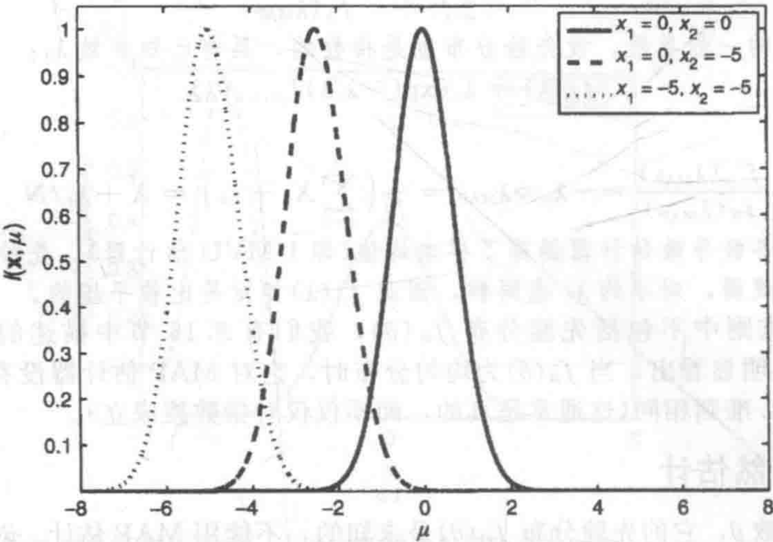


图 9-12 对 x_1 和 x_2 的特定值似然函数与参数 μ 的关系(曲线已被归一化，使得最大高度为 1)

$$L(\boldsymbol{x};\mu) = -(N/2)\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 \tag{9-262}$$

将这个表达式对 μ 微分，将结果置零，解出 $\mu_{\text{ML}}=\overline{X}$，即 μ 为样本均值。比较图 9-11 和图 9-12 中的曲线，我们看到 MAP 估计器(后验 pdf 的最大值)的结果更靠近初始值，因为估计器中用到 $1/(N+\sigma^2)$因子。当 $N\to\infty$时，MAP 和 ML 估计器相同，此时 μ 的先验分布对 MAP 估计器的影响被减弱，因为样本数非常巨大。◀

例 9-37 在例 9-36 中，当 μ 已知时，为了估计方差 σ^2，将式(9-262)对 σ^2 微分，得到：

$$\frac{\partial}{\partial\sigma^2}L(\boldsymbol{x};\sigma^2) = -\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n-\mu)^2 \tag{9-263}$$

将这个结果置零，解出 σ^2 得到：

$$\sigma^2_{\text{ML}} = \frac{1}{N}\sum_{n=1}^{N}(X_n-\mu)^2 \tag{9-264}$$

这是已知均值时的样本方差。◀

例 9-38 设 $\{X_1,\cdots,X_N\}$ 是在 $[\theta-1/2,\theta+1/2]$ 上 iid 均匀的。似然函数为：

$$\begin{aligned} l(\boldsymbol{x};\theta) &= \prod_{n=1}^{N} I_{[\theta-1/2,\theta+1/2]}(x_n) \\ &= I_{[\theta-1/2,\theta+1/2]}(x_{(1)})\,I_{[\theta-1/2,\theta+1/2]}(x_{(N)}) \end{aligned} \tag{9-265}$$

其中 $x_{(1)}$ 和 $x_{(N)}$ 是第一和第 N 阶统计量(最小和最大)。通过对样本排序，$l(\boldsymbol{x};\theta)$ 不再依赖于 $\{x_{(2)},\cdots,x_{(N-1)}\}$。正如例 9-8 中所讨论的，$\boldsymbol{T}=[X_{(1)},X_{(N)}]^{\mathrm{T}}$ 是 θ 的一个充分统计量。实际上，它是最小充分统计量，但不是完备的。注意到 $l(\boldsymbol{x};\theta)$ 不能对 θ 微分来找 ML 估计器。但是，通过对 $l(\boldsymbol{x};\theta)$ 与 θ 作图，如图 9-13b 所示，我们发现它在 $[x_{(N)}-1/2, x_{(1)}+1/2]$ 上等于 1。这可由以下式子推导得到：

$$x_{(1)} \geqslant \theta-1/2 \Rightarrow \theta \leqslant x_{(1)}+1/2 \tag{9-266}$$

$$x_{(N)} \leqslant \theta+1/2 \Rightarrow \theta \geqslant x_{(N)}-1/2 \tag{9-267}$$

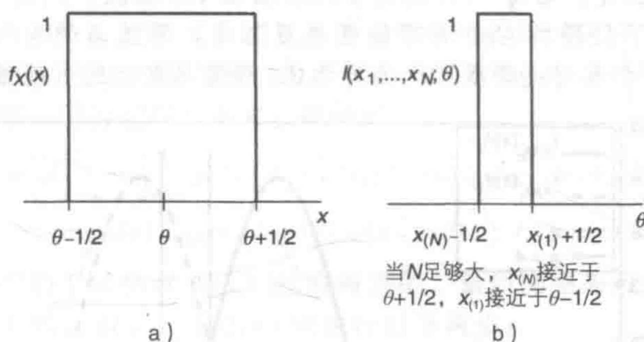


图 9-13 例 9-38 的图。a) 均匀 pdf；b) 似然函数(它不是一个关于参数 θ 的 pdf)的宽度随着 N 的增长而变窄

◀

尽管在这个结果中看起来好像将 $x_{(1)}$ 和 $x_{(N)}$ 交换了，但请注意曲线是对应的 θ。而且，随着 N 增大，我们会期望 $x_{(N)}$ 趋近于 $\theta+1/2$，$x_{(1)}$ 趋近于 $\theta-1/2$，这样图中的矩形会变得越来越窄。ML 估计对应似然函数最大，对这个例子，最大值不是唯一的：$\theta_{\text{ML}}\in[X_{(N)}-1/2, X_{(1)}+1/2]$，但它是由一阶和第 N 阶统计量定义的。

9.17 似然比检验

接下来，我们来描述一个与估计相关的基于似然函数的检测问题。在许多应用中，我们感兴趣的是：给定随机变量的一个或多个样本，确定两个模型中哪个最可能正确。例

如，设 X 为一个可测随机变量，由它确定以下模型是具有参数 θ_0 还是 θ_1。这个问题可用以下两个假设来表征：

$$H_0: X=\theta_0+V \tag{9-268}$$

$$H_1: X=\theta_1+V \tag{9-269}$$

其中包括了加性噪声随机变量 V。H_0 称为零假设，意味着它是缺省模型，我们的目标是由 X 确定 θ_1 为真，而不是 θ_0，当 $\theta_0=0$，零假设意味着这个例子中的样本只包括噪声。

定义(似然比检验) **似然比检验**(LRT)是

$$\Lambda(x)\triangleq\frac{f_{X|\Theta_1}(x|\theta_1)}{f_{X|\Theta_0}(x|\theta_0)}=\frac{l(x;\theta_1)}{l(x;\theta_0)} \tag{9-270}$$

其中条件 pdf 是具有两个不同参数值的似然函数。这也是大家熟知的纽曼—皮尔逊引理。

为了计算 LRT，需要每种假设下的似然函数，由此可以写作一个 x 的函数$\Lambda(x)$。这个结果与某个门限 η 比较，允许我们判断一种假设比另一种假设的可能性更大。如果 $\Lambda(x)>\eta$，判决假设 H_1 更有可能出现；否则，选择 H_0。如果 $\Lambda(x)=\eta$(当 X 为连续随机变量时，出现的概率为零)，检测是不确定的，我们可以选择任一种假设(例如，通过扔硬币决定)。

例 9-39 设 $\theta_0=0$，$\theta_1\neq0$ 是非随机的，假设式(9-268)中的噪声 V 是高斯的，参数为 $\{\mu=0,\sigma\}$，则 LRT 为

$$\Lambda(x)=\frac{\exp(-(x-\theta_1)^2/2\sigma^2)}{\exp(-x^2/2\sigma^2)}=\exp((2\theta_1 x-\theta_1^2)/2\sigma^2)\underset{H_0}{\overset{H_1}{\gtrless}}\eta \tag{9-271}$$

其中分子和分母中的共同项被消掉了。对其取对数，解出 x，得到了检测条件：

$$x\underset{H_0}{\overset{H_1}{\gtrless}}\frac{2\sigma^2\ln(\eta)+\theta_1^2}{2\theta_1} \tag{9-272}$$

如果假设 $\theta_1>0$，则当除以 $2\theta_1$ 时，不需要改变不等式的方向。式(9-272)的右边给出了两种判决的门限，图 9-14 画出了两个不同的 η 值。当 $\eta=1$，两种假设权值相等，判决门限在两个条件 pdf 的交叉处。当 $\eta=4$，则更多的权重在 H_0 假设，判决门限移向右边。输出 x 的样本出现在右移了门限的右边的可能性就更低了，意味着更有可能做出零假设的判决。当 $\eta>1$，要做出非零判决需要更大的样本(这种情况发生的可能性较少)。

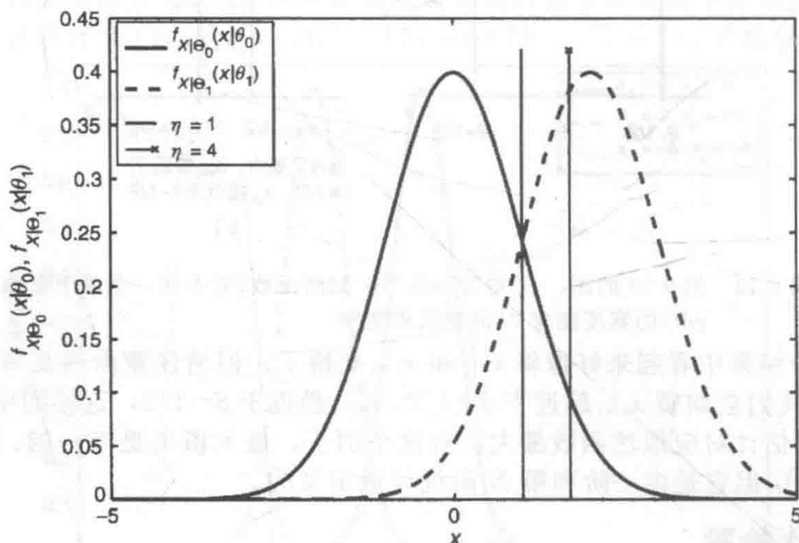


图 9-14 例 9-39 中的条件 pdf，给出了不同门限 η 的边界(竖线)

9.18　期望值最大算法

期望值最大(EM)算法是存在不可观察的因素影响样本的情况下找到 ML 估计的一种递归的方法。在传统的 ML 估计中，似然函数 $l(\boldsymbol{x};\theta)$，或等价的对数似然函数 $L(\boldsymbol{x};\theta)$ 被用来由 $\{X_1,\cdots,X_N\}$ 估计 θ。假设样本在此意义下是不完备的，还有其他与 $\{X_n\}$ 相关的随机变量。但他们不能直接测量。设这些随机变量被记为 $\boldsymbol{Y} \triangleq [Y_1,\cdots,Y_M]^{\mathrm{T}}$。

考虑可测 $\boldsymbol{X}$ 和不可测 $\boldsymbol{Y}$ 的联合 pdf $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta)$。如果 $\boldsymbol{Y}$ 也可测，则 θ 的 ML 估计器是使得联合 pdf 最大的 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的函数。但由于只有 $\boldsymbol{X}$ 可观测，有可能仅通过使 $f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ 最大来估计 θ，这可以被表示为：

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta)\mathrm{d}\boldsymbol{y} \tag{9-273}$$

基于可测的样本实现的传统的 ML 估计中不可测的 $\boldsymbol{Y}$ 对估计的影响被忽略了。这样得到的估计的方差不可能比那种 $\boldsymbol{Y}$ 可被观测而使联合 pdf $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta)$ 最大得到的估计器的方差小。EM 算法通过采用两步过程来提高 $f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ 的最大值，其中 θ 被迭代更新(M 步)，这样在每次更新之间，不可观测的 $\boldsymbol{Y}$ 的分布被估计(E 步)。EM 算法确保对 θ 的估计得到改善(或者至少没有变得更糟)，因为在第 E 步时用了一个更低的边界。

而在传统的 ML 估计中，通常使对数似然函数最大更容易些：

$$\begin{aligned}\log(f_{\boldsymbol{X}}(\boldsymbol{x};\theta)) &=\log\left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta)\mathrm{d}\boldsymbol{y}\right)\\ &=\log\left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x};\theta_{\mathrm{EM}}[k])f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta)}{f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x};\theta_{\mathrm{EM}}[k])}\mathrm{d}\boldsymbol{y}\right)\end{aligned} \tag{9-274}$$

其中基于第 k 次迭代，θ 的 EM 估计的 $f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x};\theta_{\mathrm{EM}}[k])$ 被代入分子和分母中。将附录 F 中的杰森不等式用于等式右边得到：

$$\log(f_{\boldsymbol{X}}(\boldsymbol{x};\theta))\leqslant\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x};\theta_{\mathrm{EM}}[k])\log\frac{f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta)}{f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x};\theta_{\mathrm{EM}}[k])})\mathrm{d}\boldsymbol{y} \tag{9-275}$$

则代之以原来的使 $\log(f_{\boldsymbol{X}}(\boldsymbol{x};\theta))$ 最大(只基于 $\boldsymbol{X}$)，现在只要选择 θ 使式(9-275)右边最大，而它是基于 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$。这样可由杰森不等式确保对初始问题而言上式右边是一个更低的边界，它是式(9-274)左边的最大值。由于式(9-275)比值的对数是两个对数的差，在最大化中可忽略分母项，因为它与 θ 无关。则得到：

$$\begin{aligned}&\max_{\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x};\theta_{\mathrm{EM}}[k])\log(f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta))\mathrm{d}\boldsymbol{y}\\ &=\max_{\theta}\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X};\theta_{\mathrm{EM}}[k]}[\log(f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta))]\end{aligned} \tag{9-276}$$

其中期望值是对最近的 EM 估计 $\theta_{\mathrm{EM}}[k]$运算得到的。最后的结果描述了 EM 算法：由一个初始的估计 $\theta_{\mathrm{EM}}[0]$开始，对 $k=1,2,\cdots$ 则执行以下两步：

- 求期望。计算以下条件期望：

$$\xi_{\mathrm{EM}}(\theta,\theta_{\mathrm{EM}}[k])\triangleq\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X};\theta_{\mathrm{EM}}[k]}[\log(f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y};\theta))] \tag{9-277}$$

- 求最大值。找到使以上期望值最大的 $\theta_{EM}[k+1]$：

$$\theta_{\mathrm{EM}}[k+1]=\max_{\theta}\xi_{\mathrm{EM}}(\theta,\theta_{\mathrm{EM}}[k]) \tag{9-278}$$

算法在 $\theta_{\mathrm{EM}}[k+1]$变化相对很小时停止，或经过一个预先设定的步数后停止。它很容易扩展到一个向量参数 $\boldsymbol{\theta}$。

EM 算法的一个重要应用是在第 4 章中讨论过的混合模型中的参数估计问题。考虑以下有穷混合：

$$f_X(x)=\sum_{m=1}^{M}a_m f_{Z_m}(x) \tag{9-279}$$

其中 $a_m \geqslant 0$，$\sum_{m=1}^{M} a_m = 1$。设$\{Z_m\}$分布已知，参数已知。在 EM 算法的描述中，未知参数是 $\boldsymbol{\theta} = [a_1, \dots, a_M]^{\mathrm{T}}$，样本$\{X_1, \cdots, X_N\}$可测但不完备。我们还可以在 $\boldsymbol{\theta}$ 中包括$\{Z_m\}$的任何未知参数。习题 9-28 描述了一个高斯混合态，其均值未知，也需要被估计。为了继续分析，必须定义不可测样本。设 Y_n 为一个离散随机变量，指定 Z_m 对第 n 次观测 x_n 负责。它由以下联合 pdf 描述：

$$f_{X_n, Y_n}(x_n, m; \boldsymbol{\theta}) = a_m f_{Z_m}(x_n) \tag{9-280}$$

它是式(9-279)的第 m 项。Y_n 的输出是$\{1, \cdots, M\}$，对应$\{a_1, \cdots, a_M\}$。式(9-280)对 m 求和给出了式(9-279)$x = x_n$ 的边缘 pdf。当其中有一个随机变量是离散的，如本例中的 Y_n，我们可以写为：

$$f_{X_n, Y_n}(x_n, m; \boldsymbol{\theta}) = P(Y_n = m \mid X_n = x_n, \boldsymbol{\theta}) f_{X_n}(x_n, \boldsymbol{\theta}) \tag{9-281}$$

则在 EM 算法中用到的条件分布为：

$$P(Y_n = m \mid X_n = x_n; \boldsymbol{\theta}) = \frac{a_m f_{Z_m}(x_n)}{\sum_{m=1}^{M} a_m f_{Z_m}(x_n)} \tag{9-282}$$

其中将式(9-279)($x = x_n$)和式(9-280)代入到了式(9-281)中。

(9-276)式中的期望值是离散 Y_n 的和的一个集合：

$$\sum_{m_1=1}^{M} \cdots \sum_{m_N=1}^{M} \prod_{n=1}^{N} P(Y_n = m_n \mid X_n = x_n; \boldsymbol{\theta}_{\mathrm{EM}}[k]) \log\left(\prod_{l=1}^{N} f_{X_l, Y_l}(x_l, m_l; \boldsymbol{\theta}) \right) \tag{9-283}$$

其中 m 用了下标便于在 N 次求和中区分，下式被代入：

$$f_{\boldsymbol{Y} \mid \boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta}_{\mathrm{EM}}[k]) = \prod_{n=1}^{N} P(Y_n = m_n \mid X_n = x_n; \boldsymbol{\theta}_{\mathrm{EM}}[k]) \tag{9-284}$$

$$f_{\boldsymbol{X}, \boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y}; 0) = \prod_{l=1}^{N} f_{X_l, Y_l}(x_l, m_l; \boldsymbol{\theta}) \tag{9-285}$$

由于

$$\log\left(\prod_{l=1}^{N} f_{X_l, Y_l}(x_l, m_l; \boldsymbol{\theta}) \right) = \sum_{l=1}^{N} \log(f_{X_l, Y_l}(x_l, m_l; \boldsymbol{\theta})) \tag{9-286}$$

式(9-283)中对 n 的乘积中的每一项，$N-1$ 项的和等于 1(见习题 9-21)，则式(9-283)中的表达式可简化为：

$$\sum_{l=1}^{N} \sum_{m=1}^{M} P(Y_l = m \mid X_l = x_l; \boldsymbol{\theta}_{\mathrm{EM}}[k]) \log(f_{X_l, Y_l}(x_l, m; \boldsymbol{\theta})) \tag{9-287}$$

因此不再需要 m 的下标。内部的求和是一个期望值，则对这个混合问题 EM 代价函数为：

$$\xi_{\mathrm{EM}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{\mathrm{EM}}[k]) = \sum_{n=1}^{N} \mathcal{E}_{Y_n \mid X_n; \boldsymbol{\theta}_{\mathrm{EM}}[k]} [\log(f_{X_n, Y_n}(x_n, \boldsymbol{y}_n; \boldsymbol{\theta}))] \tag{9-288}$$

其中下标由 l 改成了 n。最后，将式(9-280)和式(9-282)代入得到：

$$\begin{aligned} \xi_{\mathrm{EM}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{\mathrm{EM}}[k]) &= \sum_{n=1}^{N} \mathcal{E}_{Y_n \mid X_n; \boldsymbol{\theta}_{\mathrm{EM}}[k]} [\log(a_m f_{Z_m}(x_n))] \\ &= \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{A_m[k] f_{Z_m}(x_n)}{\sum_{m=1}^{M} A_m[k] f_{Z_m}(x_n)} \log(a_m f_{Z_m}(x_n)), \end{aligned} \tag{9-289}$$

其中 $\boldsymbol{\theta}_{\mathrm{EM}}[k] = [A_1[k], \cdots, A_M[k]]^{\mathrm{T}}$ 是 EM 估计(这是一个随机向量，因此我们用大写的 $A_M[k]$表示)。

为了使这个表达式在取某个 $\boldsymbol{\theta}$ 值时最大，我们可以写为 $\log(a_m f_{Z_m}(x_n)) = \log(a_m) + \log(f_{Z_m}(x_n))$，省略第二项，因为它与 $\boldsymbol{\theta}$ 无关。得到：

$$\xi_{\mathrm{EM}}(\boldsymbol{\theta},\boldsymbol{\theta}_{\mathrm{EM}}[k])=\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{A_m[k]f_{Z_m}(x_n)}{\sum_{m=1}^{M}A_m[k]f_{Z_m}(x_n)}\log(a_m)=\sum_{m=1}^{M}B_m[k]\log(a_m) \quad (9\text{-}290)$$

其中

$$B_m[k]\triangleq\sum_{n=1}^{N}\frac{A_m[k]f_{Z_m}(x_n)}{\sum_{m=1}^{M}A_m[k]f_{Z_m}(x_n)} \quad (9\text{-}291)$$

是一个随机系数，它依赖于当前的估计 $\boldsymbol{\theta}_{\mathrm{EM}}[k]$和不完备样本$\{X_n\}$。由于 pdf$\{f_{Z_m}(x_n)\}$已知，$\{B_m[k]\}$的值可以用每个新的样本集合和 $\boldsymbol{\theta}$ 的更新的估计算出。这使得对有穷混合的第 E 步 $\xi_{\mathrm{EM}}(\boldsymbol{\theta},\boldsymbol{\theta}_{\mathrm{EM}}[k])$ 的推导过程更完备了，其中$\{a_m\}$是待估参数。

M 步是式(9-290)的最大化，可以用拉格朗日乘子的方法来约束$\{a_m\}$。修正的代价函数为：

$$\widetilde{\xi}_{\mathrm{EM}}(\boldsymbol{\theta},\boldsymbol{\theta}_{\mathrm{EM}}[k])=\sum_{m=1}^{M}B_m[k]\log(a_m)+\lambda\Big(\sum_{m=1}^{M}a_m-1\Big) \quad (9\text{-}292)$$

其中 λ 是一个拉格朗日乘子。将其对 a_m 微分，将结果置零得到更新的估计：

$$B_m[k]/A_m[k+1]+\lambda=0\Rightarrow A_m[k+1]=-B_m[k]/\lambda \quad (9\text{-}293)$$

对 m 求和得到拉格朗日乘子：

$$\sum_{m=1}^{M}A_m[k+1]=-(1/\lambda)\sum_{m=1}^{M}B_m[k]=1\Rightarrow\lambda=-\sum_{m=1}^{M}B_m[k] \quad (9\text{-}294)$$

则 M 步的最大化是：

$$A_m[k+1]=\frac{B_m[k]}{\sum_{m=1}^{M}B_m[k]}=B_m[k]/N \quad (9\text{-}295)$$

其中最后的结果是对式(9-291)求和得到的。参数估计的结果是非负的，正如在有穷混合中要求的。如果不是这种情况，对于非负的约束，我们需要在 $\widetilde{\xi}_{\mathrm{EM}}(\boldsymbol{\theta},\boldsymbol{\theta}_{\mathrm{EM}}[k])$ 包括额外的拉格朗日乘子。

确定何时停止 EM 算法的一种方法是在每次 $\boldsymbol{\theta}_{\mathrm{EM}}[k]$更新后对对数似然函数进行评估。对上文中的混合模型，它可由下式给定：

$$L(\boldsymbol{x};\boldsymbol{\theta}_{\mathrm{EM}}[k])\log\Big(\prod_{n=1}^{N}f_{X_n}(x_n)\Big)=\sum_{n=1}^{N}\log\Big(\sum_{m=1}^{M}A_m[k]f_{Z_m}(x_n)\Big) \quad (9\text{-}296)$$

对某些小的 ε，当归一化差值满足下式时

$$D[k]\triangleq(1/N)\,|\,L(\boldsymbol{x};\boldsymbol{\theta}_{\mathrm{EM}}[k+1])-L(\boldsymbol{x};\boldsymbol{\theta}_{\mathrm{EM}}(k])\,|<\varepsilon \quad (9\text{-}297)$$

算法会停止更新。

但是，由于 EM 算法一般是非线性的，不能确保它会收敛于最佳解；它也可能收敛于一个局部最大值。

例 9-40 考虑以下 pdf：

$$f_X(x)=(0.2)\frac{1}{\sqrt{2\pi}}\exp(-(x+1)^2/2)+(0.8)\frac{1}{\sqrt{8\pi}}\exp(-(x-1)^2/8) \quad (9\text{-}298)$$

它是已知参数分别为 $\{\mu_1=-1,\sigma_1=1\}$ 和 $\{\mu_2=1,\sigma_2=2\}$ 的两个高斯 pdf 的一个有限混合，如图 9-15a 所示。由 $N=1000$ 个混合样本，我们用 EM 算法来估计 $\boldsymbol{\theta}=[a_1,a_2]^{\mathrm{T}}=[0.2,0.8]^{\mathrm{T}}$。仿真中首先得到一个在$[0,1]$上均匀分布的样本 v。如果 $v<a_1=0.2$，可由第一个高斯 pdf 导出一个样本；否则，由第二个高斯 pdf 导出一个样本。这种双重的随机性可从混合高斯中产生 1000 个样本。由初始条件 $\boldsymbol{\theta}_{\mathrm{EM}}[0]=[0.5,0.5]^{\mathrm{T}}$ 开始，EM 算法用式(9-291)和式(9-295)的方程反复更新 $\theta_{\mathrm{EM}}[k]$。图 9-15b 显示了轨迹(现实)，我们看到 EM 算法很快收敛到正确的权值。差值 $D[k]$的一个现实也显示在图中，最终估计的混合由图 9-15b 中给出。

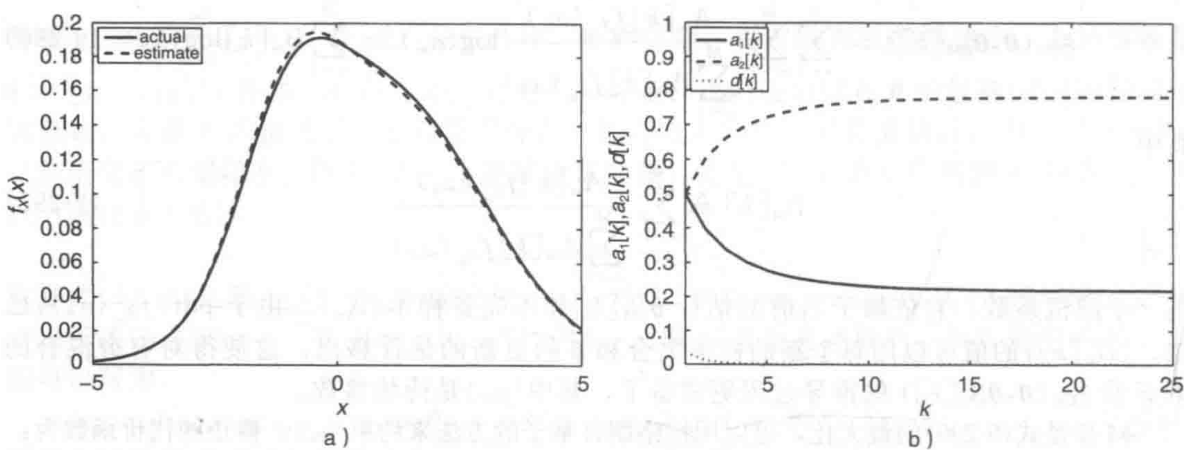


图 9-15 例 9-40 中对有穷混合的 EM 算法的结果。a) 实际的和估计的高斯混合和估计；b) 参数 $\{A_1[k],A_2[k]\}$ 的收敛性和对数似然性能测度 $D[k]$（它非常接近于零）

在第 12 章，我们会检测其他类型的自适应算法来找到使代价函数最小的方法，比如 MSE。

9.19 矩方法

矩方法(MoM)是一种参数估计技术，等同于样本矩，它与输出和总体矩相关，依赖于分布参数。由第 5 章可知随机变量 X 的 m 阶原点矩为：

$$\mu_m \triangleq \mathcal{E}[X^m] \tag{9-299}$$

设相应的样本矩为：

$$\overline{X^m} \triangleq \frac{1}{N}\sum_{n=1}^{N} X_n^m \tag{9-300}$$

对任意 pdf $f_X(x)$，我们知道 $\mu_1=\mu_X$，$\mu_2=\sigma_X^2-\mu_X^2$，其中 $\{\mu_X,\sigma_X^2\}$ 是分布的集合均值和方差。对 MoM 估计，用以下方程解出 μ_X 和 σ_X^2：

$$\overline{X}=\mu_1=\mu_X,\quad \overline{X^2}=\mu_2=\sigma_X^2+\mu_X^2 \tag{9-301}$$

它生成了以下估计：

$$\mu_{\text{MoM}}=\overline{X},\quad \sigma_{\text{MoM}}^2=\overline{X^2}-\overline{X}^2=[(N-1)/N]S^2 \tag{9-302}$$

其中 S^2 是样本方差。对高斯随机变量，μ_X 和 σ_X^2 明确地出现在 pdf 中，因此均值和方差的估计可直接由式(9-302)得到。结果显示 ML 估计与式(9-302)得到的结果相同，对于其他分布情况通常不同。

例 9-41 考虑伽马随机变量，参数为 $\{\alpha,r\}$。由附录 A：

$$\mathcal{E}[X]=r\alpha,\quad \mathcal{E}[X^2]=r(r+1)\alpha^2 \tag{9-303}$$

将其等于样本矩，解出参数得到：

$$\alpha_{\text{MoM}}=\frac{\overline{X^2}-\overline{X}^2}{\overline{X}},\quad r_{\text{MoM}}=\frac{\overline{X}^2}{\overline{X^2}-\overline{X}^2} \tag{9-304}$$

对 ML 估计，似然函数为：

$$l(\boldsymbol{x};\alpha,r)=\frac{1}{\alpha^{Nr}\Gamma^N(r)}\exp\Big(-\sum_{n=1}^{N}x_n/\alpha\Big)\prod_{n=1}^{N}x_n^{r-1} \tag{9-305}$$

对数似然函数为：

$$L(\boldsymbol{x};\alpha,r)=-N(r\ln(\alpha)+\Gamma(r)]-(1/\alpha)\sum_{n=1}^{N}x_n+(r-1)\sum_{n-1}^{N}\ln(x_n) \tag{9-306}$$

将这个表达式对每个参数 $\{\alpha,r\}$ 微分，将结果置零，我们看到不可能得到一个 r 的估计的闭集表达式，因此我们只关注 α：

$$\frac{\partial}{\partial\alpha}L(\boldsymbol{x};\alpha,r)=-Nr/\alpha+(N/\alpha^2)\overline{X}=0 \tag{9-307}$$

由上式我们得到：

$$\alpha_{\mathrm{ML}}=\overline{X}/r \tag{9-308}$$

这是已知 r 时的 ML 估计器，也是在已知 r 时 α 的 MoM 估计器的结果。◀

9.20 最小二乘估计

假设参数向量 $\boldsymbol{\theta}\in\mathcal{R}^M$ 的信息是通过以下的测量模型得到的：

$$Y=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\theta}+V \tag{9-309}$$

其中 Y 是一个随机变量，$\boldsymbol{a}$ 是一个已知向量，V 是一个无法观测的加性噪声随机变量。这是之前讨论过的简单模型 $Y=X+V$ 的扩展，X 是一个随机变量。如果 $\boldsymbol{\theta}$ 为一个随机向量，则在最小二乘(LS)估计中不需考虑它的分布。假设有 N 个 iid 样本 $\boldsymbol{Y}\triangleq[Y_1,\cdots,Y_N]^{\mathrm{T}}$，则

$$\boldsymbol{Y}=\boldsymbol{A\theta}+\boldsymbol{V} \tag{9-310}$$

其中 $\boldsymbol{V}\triangleq[V_1,\cdots,V_N]^{\mathrm{T}}$。

$$\boldsymbol{A}\triangleq[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_N]^{\mathrm{T}}\in\mathcal{R}^{N\times M} \tag{9-311}$$

通常，样本的数目要远超参数的数目，即 $N\gg M$，则 $\boldsymbol{A}$ 是一个又高又窄的矩阵。设 $\hat{\boldsymbol{Y}}\triangleq\boldsymbol{A}\,\hat{\boldsymbol{\theta}}$是基于参数估计$\hat{\boldsymbol{\theta}}$的测量的估计，定义估计误差为$\widetilde{\boldsymbol{Y}}=\boldsymbol{Y}-\hat{\boldsymbol{Y}}$

定义(加权 LS 估计器)　参数向量 θ 的**加权 LS 估计器**(WLS)为：

$$\boldsymbol{\theta}_{\mathrm{WLS}}=\min_{\theta}\widetilde{\boldsymbol{Y}}^{\mathrm{T}}\boldsymbol{W}\,\widetilde{\boldsymbol{Y}} \tag{9-312}$$

其中 $\boldsymbol{W}\in\mathcal{R}^{N\times N}$ 为加权矩阵。

加权矩阵的目的是更重视某些样本。例如，如果 $\boldsymbol{\theta}$ 正好是随时间而改变，那么对更近的样本更重视则显然更有利。

代价函数为：

$$\xi_{\mathrm{WLS}}\triangleq\widetilde{\boldsymbol{Y}}^{\mathrm{T}}\boldsymbol{W}\,\widetilde{\boldsymbol{Y}}=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{WY}-2\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{WA\theta}+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WA\theta} \tag{9-313}$$

它是 $\boldsymbol{\theta}$ 的二次函数。通过对 $\boldsymbol{\theta}$ 微分并将结果置为零向量很容易得到最小值，

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WA\theta}_{\mathrm{WLS}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WY} \tag{9-314}$$

其中 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WA}\in\mathcal{R}^{M\times M}$，$\boldsymbol{WA}\in\mathcal{R}^{N\times M}$。如果 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WA}$ 是满秩的($=M$)，则它是可逆的，因此估计器为：

$$\boldsymbol{\theta}_{\mathrm{WLS}}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WA})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WY} \tag{9-315}$$

当加权矩阵为单位矩阵即 $\boldsymbol{W}=\boldsymbol{I}$ 时，得到传统 LS 估计。

例 9-42　考虑最简单的情况，即 $\boldsymbol{\theta}\to\theta$ 为一个标量($M=1$)，测量模型为

$$Y=\theta+V \tag{9-316}$$

其中 V 具有零均值。假设 N 次测量，$\boldsymbol{W}=\boldsymbol{I}$，变量为

$$\boldsymbol{a}=1,\boldsymbol{A}=[1,\cdots,1]^{\mathrm{T}},\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=N,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y}=\sum_{n=1}^{N}Y_n \tag{9-317}$$

则 LS 估计器为样本均值：$\theta_{\mathrm{LS}}=\overline{Y}$ ◀

例 9-43　接下来，设测量模型稍微复杂些

$$Y=X\theta+V \tag{9-318}$$

其中参数仍为一个标量，但其乘以了一个与 V 不相关的零均值随机变量 X。假设 N 次测量，$\boldsymbol{a}_n=X_n$

$$\boldsymbol{A}=[X_1,\cdots,X_N]^{\mathrm{T}},\boldsymbol{A}^{T}\boldsymbol{A}=\sum_{n=1}^{N}X_n^2,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y}=\sum_{n=1}^{N}X_nY_n \tag{9-319}$$

它给出了以下的 LS 估计器：

$$\theta_{LS} = \frac{\sum_{n=1}^{N} X_n Y_n}{\sum_{n=1}^{N} X_n^2} = \frac{1}{N\,\overline{X^2}} \sum_{n=1}^{N} X_n Y_n \tag{9-320}$$

这个结果就是大家熟知的在线性回归中得到的准则，即如何确定 N 个样本对 $\{X_n, Y_n\}$ 的最佳线性拟合的准则。分子是确定的两个样本集的互相关，分母是 X 的样本的二阶矩。图 9-16 显示的是$N=100$时这个结论的一个示例，其中$\{X_n\}$是 iid 标准高斯分布。噪声 V 也是零均值的 iid 高斯型，但方差稍小，为 0.5。未知参数是 $\theta=2$，图中的直线是$\hat{Y}=\theta_{LS}X$，其斜率为 $\theta_{LS}=1.9219$，因此给出了样本的正确的拟合。观察到直线正如所料穿过了原点，因为模型中并没有包括纵坐标项，如例 9-44 中所描述的。

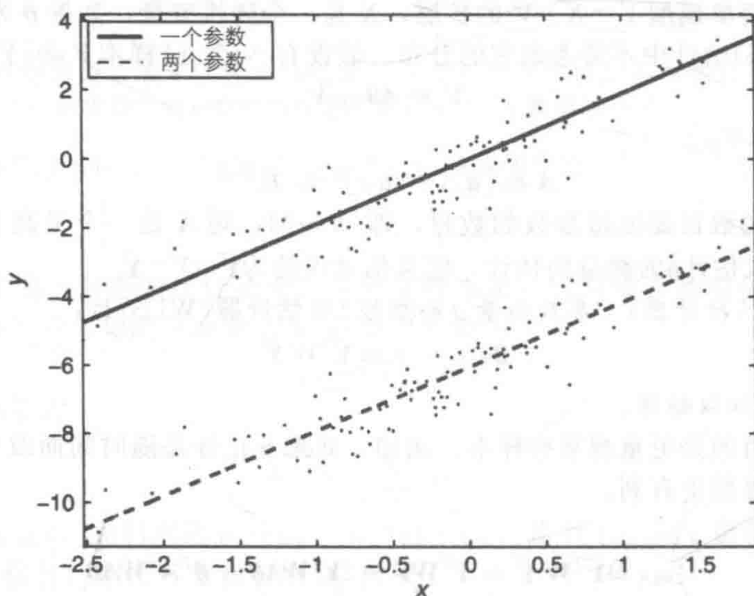


图 9-16　例 9-43 和例 9-44 的线性回归模型

例 9-44 将之前的模型扩展为含有两个参数，我们得到：

$$Y = \theta_1 + \theta_2 X + V \tag{9-321}$$

则 $\boldsymbol{a}_n^{\mathrm{T}} = [1, X_n]$，

$$\boldsymbol{A} = \begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_N \end{bmatrix}^{\mathrm{T}}, \quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y} = \begin{bmatrix} \sum_{n=1}^{N} Y_n \\ \sum_{n=1}^{N} X_n Y_n \end{bmatrix} = N \begin{bmatrix} \overline{Y} \\ (1/N)\sum_{n=1}^{N} X_n Y_n \end{bmatrix} \tag{9-322}$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} N & \sum_{n=1}^{N} X_n \\ \sum_{n=1}^{N} X_n & \sum_{n=1}^{N} X_n^2 \end{bmatrix} = N \begin{bmatrix} 1 & \overline{X} \\ \overline{X} & \overline{X^2} \end{bmatrix} \tag{9-323}$$

两个 LS 估计为：

$$\theta_{LS,1} = \frac{1}{\Delta}\left[\overline{X^2}\,\overline{Y} - (1/N)\overline{X}\sum_{n=1}^{N} X_n Y_n\right] \tag{9-324}$$

$$\theta_{LS,2} = \frac{1}{\Delta}\left[(1/N)\sum_{n=1}^{N} X_n Y_n - \overline{X}\,\overline{Y}\right] \tag{9-325}$$

其中
$$\Delta \triangleq \overline{X^2} - \overline{X}^2 \tag{9-326}$$
为$(1/N)\boldsymbol{A}^T\boldsymbol{A}$ 的行列式。对应 $\theta_1=-6$ 和 $\theta_2=2$ 的这种情况的仿真也在图 9-16 中显示(标上了两个参数)，用了与例 9-43 中一样的 $\{X_n, Y_n\}$ 样本。直线由下式给定：
$$\hat{Y} = \theta_{LS,1} + \theta_{LS,2} X \tag{9-327}$$
它不经过原点：估计值 $\theta_{LS,2}=1.9210$ 和 $\theta_{LS,1}=-6.0252$ 非常接近模型值。最后，由于 X 的均值为零，我们可以预期对于大的 N，这两个参数估计的第二项，会如 Δ 的第二项一样近似为零。将 $\overline{X}\approx 0$ 代入，估计简化为：
$$\theta_{LS,1} \approx \overline{Y}, \quad \theta_{LS,2} \approx \frac{1}{N\overline{X^2}}\sum_{n=1}^{N} X_n Y_n \tag{9-328}$$

斜率 $\theta_{LS,2}$ 与例 9-43 中一样，纵坐标 $\theta_{LS,1}$ 是 Y 的样本均值，这正是预期的结果，因为模型中的 X 和 V 都是零均值。◀

对多重线性回归，目标是找到以下模型的参数 $\{\theta_1,\cdots,\theta_M\}$
$$Y = \sum_{m=1}^{M} \theta_m X_m + V \tag{9-329}$$
这个模型在我们看来是例 9-44 中模型的扩展，包括了几个随机变量 $\{X_1,\cdots,X_M\}$。

对 N 个样本，它与之前式(9-310)中描述的模型相似
$$\boldsymbol{A} = \begin{bmatrix} X_{11} & \cdots & X_{1M} \\ \vdots & \cdots & \vdots \\ X_{N1} & \cdots & X_{NM} \end{bmatrix} \tag{9-330}$$
其中 $\{X_{1m},\cdots,X_{Nm}\}$ 是第 m 个随机变量 X_m 的 N 个样本。

例 9-45 考虑以下包括 Hald 数据(Hald，1952)的矩阵(由 MATLAB 得到)，其中 $N=13$ 个测量值，$M=4$ 个变量
$$\boldsymbol{A}^T = \begin{bmatrix} 7 & 1 & 11 & 11 & 7 & 11 & 3 & 1 & 2 & 21 & 1 & 11 & 10 \\ 26 & 29 & 56 & 31 & 52 & 55 & 71 & 31 & 54 & 47 & 40 & 66 & 68 \\ 6 & 15 & 8 & 8 & 6 & 9 & 17 & 22 & 18 & 4 & 23 & 9 & 8 \\ 60 & 52 & 20 & 47 & 33 & 22 & 6 & 44 & 22 & 26 & 34 & 12 & 12 \end{bmatrix} \tag{9-331}$$

对应 $\boldsymbol{A}$ 的 Hald 数据的输出的测量值为：
$$\boldsymbol{Y} = (78.5, 74.3, 104.3, 87.6, 95.9, 109.2, 102.7, 72.5, 93.1, 115.9, 83.8, 113.3, 109.4]^T \tag{9-332}$$

因此
$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1139 & 4922 & 769 & 2620 \\ 4922 & 33050 & 7201 & 15739 \\ 769 & 7201 & 2293 & 4628 \\ 2620 & 15739 & 4628 & 15062 \end{bmatrix}, \quad \boldsymbol{A}^T\boldsymbol{Y} = [10032, 62028, 13982, 34733]^T \tag{9-333}$$
$$\boldsymbol{\theta}_{LS} = [2.1930, 1.1533, 0.7585, 0.4863]^T$$
◀

这一节我们通过对主分量分析(这个技术可用来产生一个降阶的 LS 估计)的简要讨论得到一些结论。在关于线性回归和 LS 估计中，它被称为*主成分回归*(PCR)。由于随机变量 $\{X_1,\cdots,X_M\}$ 可能不是零均值，有必要将样本均值从 $\boldsymbol{A}$ 的元素中减掉来计算在 PCR 中用到的样本的协方差阵。样本均值可以表示为下式：
$$\overline{\boldsymbol{X}} = (1/N)\boldsymbol{A}^T\boldsymbol{1} \tag{9-334}$$
其中 $\boldsymbol{1}\in\mathcal{R}^N$ 是一个全 1 向量，比例因子 $1/N$ 使得 $\overline{\boldsymbol{X}}$ 无偏。则由 $\overline{\boldsymbol{X}}$ 用克罗内克积构成一个矩阵(见附录 G)：
$$\boldsymbol{M} = \boldsymbol{1} \otimes \overline{\boldsymbol{X}}^T \tag{9-335}$$
$\boldsymbol{M}$ 的第一列是重复 N 次的 X_1 的样本均值，$\boldsymbol{M}$ 其他列和 X_n 也同样。因此，样本的协方差

阵为

$$\hat{C}_{XX} = \frac{1}{N-1}(\boldsymbol{A}-\boldsymbol{M})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{M}) \tag{9-336}$$

其中包括了比例因子 $1/(N-1)$，这样它是无偏的。对$\hat{C}_{XX}$特征分解得到：

$$\hat{C}_{XX} = \hat{Q}\Lambda Q^{-1} \tag{9-337}$$

其中 $\boldsymbol{\Lambda}$ 包括$\hat{C}_{XX}$特征值，$\boldsymbol{Q}$ 的列是相应的特征向量。

接下来，我们将初始数据变换为 $\boldsymbol{B}\triangleq\boldsymbol{AQ}$，将式(9-310)中模型重写为如下：

$$\boldsymbol{Y} = \boldsymbol{AQQ}^{-1}\boldsymbol{\theta} + \boldsymbol{V} = \boldsymbol{B}\boldsymbol{\theta}_{\mathrm{PCA}} + \boldsymbol{V} \tag{9-338}$$

其中 $\boldsymbol{\theta}_{\mathrm{PCA}}\triangleq\boldsymbol{Q}^{-1}\boldsymbol{\theta}$。这个修正模型的 LS 估计为：

$$\boldsymbol{\theta}_{\mathrm{PCA,LS}} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y} \tag{9-339}$$

它与原始的 LS 解相同，因为

$$\begin{aligned}\boldsymbol{Q}\boldsymbol{\theta}_{\mathrm{PCA,LS}} &= \boldsymbol{Q}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AQ})^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y} \\ &= (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y} = \boldsymbol{\theta}_{\mathrm{LS}}\end{aligned} \tag{9-340}$$

由特征分解，我们可决定保留与 $\boldsymbol{\Lambda}$ 的 $L<M$ 最大的特征值对应的特征向量。设矩阵为 $\boldsymbol{Q}_{\mathrm{PCA}}\in\mathcal{R}^{M\times L}$，定义 $\boldsymbol{B}_{\mathrm{PCA}}\triangleq\boldsymbol{AQ}_{\mathrm{PCA}}\in\mathcal{R}^{N\times L}$，得到降阶的模型：

$$\boldsymbol{Y} = \boldsymbol{B}_{\mathrm{PCA}}\boldsymbol{\theta}_{\mathrm{PCA}} + \boldsymbol{V} \tag{9-341}$$

当前 $\boldsymbol{\theta}_{\mathrm{PCA}}\in\mathcal{R}^{L}$。LS 解为：

$$\boldsymbol{\theta}_{\mathrm{LS,PCA}} = (\boldsymbol{B}_{\mathrm{PCA}}^{\mathrm{T}}\boldsymbol{B}_{\mathrm{PCA}})^{-1}\boldsymbol{B}_{\mathrm{PCA}}\boldsymbol{Y} \tag{9-342}$$

测量向量的估计为$\hat{\boldsymbol{Y}}=\boldsymbol{B}_{\mathrm{PCA}}\boldsymbol{\theta}_{\mathrm{LS,PCA}}$。但我们还应该注意到与小的特征值对应的特征向量也是重要的，这取决于 $\boldsymbol{A}$ 和 $\boldsymbol{Y}$，因此忽略这些分量并不总是会带来好处(Jolliffe，1982)。

例 9-46 对例 9-45 的 Hald 数据，样本均值为：

$$\overline{X} = [7.4615, 48.1538, 11.7692, 30.0000]^{\mathrm{T}} \tag{9-343}$$

样本的协方差阵为：

$$\hat{C}_{XX} = \begin{bmatrix} 34.6026 & 20.9231 & -31.0513 & -24.1667 \\ 20.9231 & 242.1410 & -13.8782 & -253.4167 \\ -31.0513 & -13.8782 & 41.0256 & 3.1667 \\ -24.1667 & -253.4167 & 3.1667 & 280.1667 \end{bmatrix} \tag{9-344}$$

对$\hat{C}_{XX}$做特征分解得到以下特征向量矩阵：

$$\boldsymbol{Q} = \begin{bmatrix} -0.0678 & 0.6460 & -0.5673 & 0.5062 \\ -0.6785 & 0.0200 & 0.5440 & 0.4933 \\ 0.0290 & -0.7553 & -0.4036 & 0.5156 \\ 0.7309 & 0.1085 & 0.4684 & 0.4844 \end{bmatrix} \tag{9-345}$$

特征值矩阵 $\boldsymbol{\Lambda}$ 具有对角元素{517.7969，67.4964，12.4054，0.2372}。由于特征值范围≈2183，我们可以忽略与特征值 0.2372 对应的第四个特征向量。只为 $\boldsymbol{Q}_{\mathrm{PCA}}$ 保留 $\boldsymbol{Q}$ 的前三列得到：

$$\boldsymbol{B}_{\mathrm{PCA}}^{\mathrm{T}}\boldsymbol{B}_{\mathrm{PCA}} = \begin{bmatrix} 7761 & -21 & -4435 \\ -21 & 810 & 60 \\ -4435 & 60 & 12855 \end{bmatrix},\quad \boldsymbol{B}_{\mathrm{PCA}}^{\mathrm{T}}\boldsymbol{Y} = [-16976929, 38677]^{\mathrm{T}} \tag{9-346}$$

$\boldsymbol{\theta}_{\mathrm{LS,PCA}}=[-0.5830，0.9227，2.8032]^{\mathrm{T}}$。◀

9.21 LS 估计器的性质

在这节中，我们给出式(9-315)中的加权 LS 估计器的某些特性，用到式(9-311)中的矩阵 $\boldsymbol{A}$。

9.21.1 最小 ξ_{WLS}

将 WLS 估计器代入式(9-313)中得到：

$$\begin{aligned}\xi_{\text{WLS,min}} &= \boldsymbol{Y}^{\text{T}}\boldsymbol{W}\boldsymbol{Y} - 2\boldsymbol{Y}^{\text{T}}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{Y} \\ &\quad + \boldsymbol{Y}^{\text{T}}\boldsymbol{W}^{\text{T}}\boldsymbol{A}(\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{Y} \\ &= \boldsymbol{Y}^{\text{T}}\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{Y}^{\text{T}}\boldsymbol{W}\boldsymbol{A}(\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{Y} \\ &= \boldsymbol{Y}^{\text{T}}\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{Y}^{\text{T}}\boldsymbol{W}\boldsymbol{A}\boldsymbol{\theta}_{\text{WLS}}\end{aligned} \tag{9-347}$$

其中假设 $\boldsymbol{W}=\boldsymbol{W}^{\text{T}}$ 为对称的。代入 LS 参数导致扩展式中的第三项被消去，没有进一步简化的可能后在最后一个表达式中插入 $\boldsymbol{\theta}_{\text{WLS}}$项。例 9-44 中 $\boldsymbol{W}=\boldsymbol{I}$ 时的 LS 代价函数 ξ_{LS}的曲线在图 9-17a 中给出。它具有凸的二次型，最小值靠近 $\boldsymbol{\theta}=[2,-6]^{\text{T}}$。注意到 ξ_{WLS}不是一个集合值；它是一个样本的函数，因此曲线会随着 X 和 Y 不同的输出而改变。但是，对于大的 N，最小值的基本形状和位置在图中给出了。

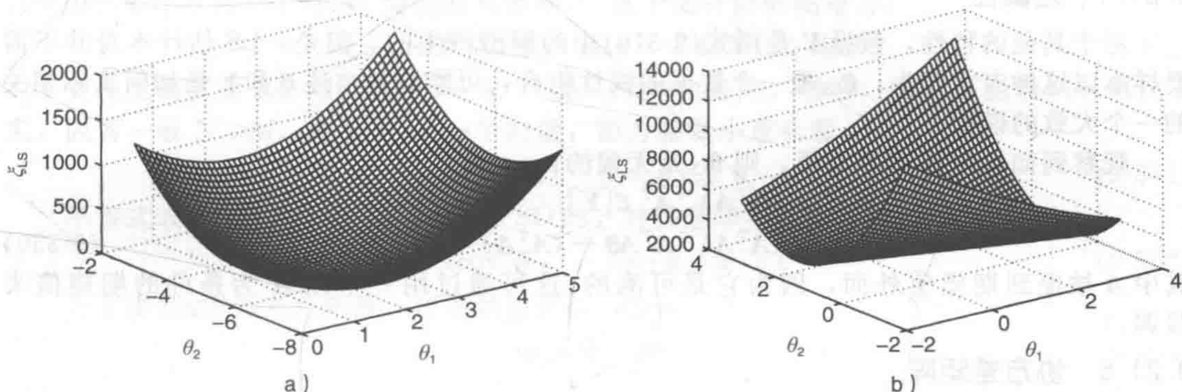


图 9-17　例 9-44 的 LS 代价函数。a) 非奇异 $\boldsymbol{A}^{\text{T}}\boldsymbol{A}$；b) 奇异 $\boldsymbol{A}^{\text{T}}\boldsymbol{A}$，因此不存在唯一的最小值

9.21.2　唯一性

LS 解答是唯一的，如果 $\boldsymbol{A}$ 是满秩的，ξ_{WLS}有唯一的最小值，意味着它有 M 个线性独立的列。当这种情况发生时(类似于 $N>>M$)，$\boldsymbol{A}^{\text{T}}\boldsymbol{A}$ 是非奇异的，它的逆存在。图 9-17b 给出了当 $\boldsymbol{A}$ 不是满秩时代价函数的一个例子；它不再是二次型并且沿着图中的凹槽具有无穷个解。这个例子是人为改变例 9-44 中的 $\boldsymbol{A}$ 来创造的，因此它的第二列是第一列的三倍。这导致了图 9-17b 的“碗形”如图所示出现了坍塌。即使图 9-17b 中最小 ξ_{WLS}对应的任意点都是一个解，对应的 $\boldsymbol{\theta}_{\text{LS}}$也不再是 $\boldsymbol{\theta}$ 的一个好的估计，得到的直线不能正确的对数据建模。

9.21.3　正交性

重写 WLS 参数的表达式得到：

$$\boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{A}^{\text{T}}\boldsymbol{W}\boldsymbol{A}\boldsymbol{\theta}_{\text{WLS}} = \boldsymbol{A}^{\text{T}}\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{\theta}_{\text{WLS}}) = \boldsymbol{A}^{\text{T}}\boldsymbol{W}\widetilde{\boldsymbol{Y}}_{\text{WLS}} = \boldsymbol{0} \tag{9-348}$$

其中$\widetilde{\boldsymbol{Y}}_{\text{LS}}$是在 WLS 解答中评估的估计误差向量，$\boldsymbol{0}\in\mathcal{R}^{M}$ 是一个零向量。这个表达式证明对 WLS 参数向量，$\boldsymbol{A}$(被 $\boldsymbol{W}$ 加权的)的列与估计误差向量正交。如果这个表达式的左边乘上 $\boldsymbol{\theta}_{\text{WLS}}$，则

$$\boldsymbol{\theta}_{\text{WLS}}^{\text{T}}\boldsymbol{A}^{\text{T}}\boldsymbol{W}\widetilde{\boldsymbol{Y}}_{\text{WLS}} = \hat{\boldsymbol{Y}}^{\text{T}}\boldsymbol{W}\widetilde{\boldsymbol{Y}}_{\text{WLS}} = 0 \tag{9-349}$$

因此 WLS 估计器和估计误差向量也是正交的。对应 $\boldsymbol{W}=\boldsymbol{I}$ 的这个关系在图 9-18 中描述了。$\boldsymbol{Y}$ 不在 $\boldsymbol{A}$ 的列空间(范围)中；否则，它将被很好地估计出。估计的输出$\hat{\boldsymbol{Y}}_{\text{LS}}$必定在 $\boldsymbol{A}$ 的列空间中，因为它是通过 $\boldsymbol{\theta}_{\text{LS}}$对 $\boldsymbol{A}$ 的列加权求和得到的。显然，$\hat{\boldsymbol{Y}}$可以是 $\boldsymbol{A}$ 范围内任意向量，但是图说明了最好的向量是与误差 $\widetilde{\boldsymbol{Y}}_{\text{LS}}$ 正交的，因为$\widetilde{\boldsymbol{Y}}_{\text{LS}}^{\text{T}}\widetilde{\boldsymbol{Y}}_{\text{LS}}$(向量模平方)是最小的。图 9-19显示的是 $N=2$ 个样本和 $M=1$ 个参数时特定的例子(这样 $\boldsymbol{A}$ 是一个列向量)，证明向量是正交的，即使只有少量的样本；正交原理是 LS 估计的一个确定特性，不只对应大样本。

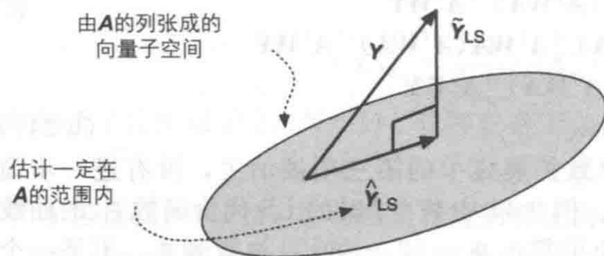


图 9-18 LS正交条件的向量空间解释

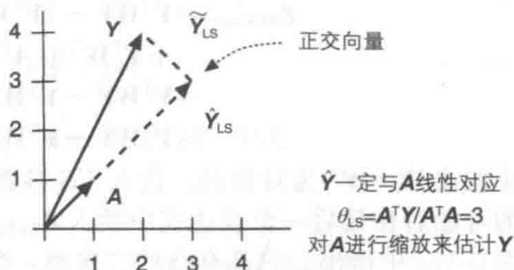


图 9-19 二维空间中LS正交条件的数字示例

9.21.4 无偏性

对于其他的特性，假设 $\boldsymbol{Y}$ 是用式(9-310)中的模型产生的。但是，LS估计本身并不需要样本以这种方式产生；$\boldsymbol{\theta}_{\text{LS}}$是一个数据的线性组合，可能只是描述 $\boldsymbol{\theta}$ 和 $\boldsymbol{Y}$ 是如何实际相关的一个大致的模型。

观察到如果 $\boldsymbol{V}$ 具有零均值，则 $\boldsymbol{\theta}_{\text{LS}}$是无偏的：

$$\begin{aligned}\mathcal{E}[\boldsymbol{\theta}_{\text{LS}}] &= (\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\,\mathcal{E}[\boldsymbol{Y}] \\ &= (\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{A}\boldsymbol{\theta} + (\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\,\mathcal{E}[\boldsymbol{V}] = \boldsymbol{\theta}\end{aligned} \tag{9-350}$$

其中 $\boldsymbol{A}$ 被提到期望值外面，因为它是可测的(这可通过用一个以 $\boldsymbol{A}$ 为条件的期望值来强调。)

9.21.5 协方差矩阵

由上式的第二行，即在求期望前，基于式(9-310)的模型的LS估计器为：

$$\boldsymbol{\theta}_{\text{LS}} = \boldsymbol{\theta} + (\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{V} \tag{9-351}$$

减去 $\boldsymbol{\theta}$(它为均值，因为估计器是无偏的)，我们可将协方差矩阵写做：

$$\begin{aligned}\mathcal{E}[(\boldsymbol{\theta}_{\text{LS}}-\boldsymbol{\theta})(\boldsymbol{\theta}_{\text{LS}}-\boldsymbol{\theta})^{\text{T}}] &= \mathcal{E}[(\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{V}\boldsymbol{V}^{\text{T}}\boldsymbol{A}(\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}) \\ &= (\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\,\mathcal{E}[\boldsymbol{V}\boldsymbol{V}^{\text{T}}]\boldsymbol{A}(\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1} \\ &= (\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\text{T}}\boldsymbol{R}_{\boldsymbol{VV}}\boldsymbol{A}(\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1}\end{aligned} \tag{9-352}$$

为继续分析问题，假设 $\boldsymbol{V}$ 是不相关的，则 $\boldsymbol{R}_{\boldsymbol{VV}}=\sigma_V^2\boldsymbol{I}$，

$$\text{cov}[\boldsymbol{\theta}_{\text{LS}}] = \sigma_V^2(\boldsymbol{A}^{\text{T}}\boldsymbol{A})^{-1} \tag{9-353}$$

这个结果不仅如预测的依赖于噪声方差，还依赖于确定的自相关矩阵 $\boldsymbol{A}^{\text{T}}\boldsymbol{A}$。

9.21.6 有效性：得到CRLB

LS估计器在假设 $\boldsymbol{V}$ 是零均值、高斯的和不相关条件下得到CRLB。LS模型的Fisher信息矩阵是

$$\mathcal{I}(\boldsymbol{\theta}) \triangleq \mathcal{E}\left[\frac{\partial \ln f_{\boldsymbol{V}}(\boldsymbol{V})}{\partial \boldsymbol{\theta}}\left(\frac{\partial \ln f_{\boldsymbol{V}}(\boldsymbol{V})}{\partial \boldsymbol{\theta}}\right)^{\text{T}}\right] \tag{9-354}$$

其中 $f_{\boldsymbol{V}}(\boldsymbol{V})$是 $\boldsymbol{V}$ 的联合pdf，$\mathcal{I}(\boldsymbol{\theta})\in\mathcal{R}^{M\times M}$，$M$ 是 $\boldsymbol{\theta}$ 中参数的数目。对于以上 $\boldsymbol{R}_{\boldsymbol{VV}}=\sigma_V^2\boldsymbol{I}$ 的假设，它可以直接计算对数：

$$\ln f_{\boldsymbol{V}}(\boldsymbol{V}) = \ln\left(\prod_{n=1}^{N} f_{V_n}(V_n)\right) = \ln\left(\frac{1}{(2\pi\sigma_V^2)^{N/2}}\right) - \frac{1}{2\sigma_V^2}\sum_{n=1}^{N} V_n^2 \tag{9-355}$$

其中第一个等式用到了独立性，第二个等式用到了高斯假设。因此

$$\frac{\partial}{\partial \boldsymbol{\theta}}\ln f_{\boldsymbol{V}}(\boldsymbol{V}) = -\frac{1}{2\sigma_V^2}\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{V}^{\text{T}}\boldsymbol{V} = -\frac{1}{\sigma_V^2}\left[\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\theta}}\right]^{\text{T}}\boldsymbol{V} \tag{9-356}$$

其中求和被求内积取代，用到了附录G中的向量微分特性。由于 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{\theta}+\boldsymbol{V}$，求导结果是：

$$\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\theta}} = \frac{\partial(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\boldsymbol{A}^{\text{T}} \tag{9-357}$$

将这个结果代入求期望值中可得：

$$\mathcal{I}(\boldsymbol{\theta})=\frac{1}{\sigma_V^4}\,\mathcal{E}\left[\left[\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\theta}}\right]^{\mathrm{T}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\,\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\theta}}\right]=\frac{1}{\sigma_V^4}\boldsymbol{A}\,\mathcal{E}[\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}]\boldsymbol{A}^{\mathrm{T}}=\frac{1}{\sigma_V^2}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} \tag{9-358}$$

显示 LS 估计器的协方差阵达到 CRLB 边界$\mathcal{I}^{-1}(\boldsymbol{\theta})=\sigma_V^2(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}$。高斯白噪声 $\boldsymbol{V}$ 的 LS 估计器是最优线性无偏(BLU)估计器，意味着它在所有无偏线性估计器中方差最小，接下来将给出证明。

9.21.7　BLU 估计器

设$\hat{\boldsymbol{\theta}}$为 $\boldsymbol{\theta}$ 的任意无偏估计器，它可被写为 $\boldsymbol{Y}$ 的线性组合，如下所示：

$$\hat{\boldsymbol{\theta}}=\boldsymbol{B}\boldsymbol{Y} \tag{9-359}$$

其中 $\boldsymbol{B}\in\mathcal{R}^{M\times N}$(为一个与 $\boldsymbol{A}^{\mathrm{T}}$ 类似的宽矩阵)。这个估计器的期望为：

$$\mathcal{E}[\hat{\boldsymbol{\theta}}]=\boldsymbol{B}\,\mathcal{E}[\boldsymbol{Y}]=\boldsymbol{B}\boldsymbol{A}\boldsymbol{\theta}+\boldsymbol{B}\,\mathcal{E}[\boldsymbol{V}]=\boldsymbol{B}\boldsymbol{A}\boldsymbol{\theta} \tag{9-360}$$

为了实现无偏需要 $\boldsymbol{B}\boldsymbol{A}=\boldsymbol{I}$，其中 $\boldsymbol{I}\in\mathcal{R}^{M\times M}$是一个单位矩阵。通常，我们不能确切的解出 $\boldsymbol{B}$ 来，因为一般 $N\gg M$。由于 $\boldsymbol{\theta}_{\mathrm{LS}}$是一个向量，协方差最小意味着：

$$\operatorname{cov}[\boldsymbol{\theta}_{\mathrm{LS}}]-\operatorname{cov}[\hat{\boldsymbol{\theta}}]\geqslant\mathbf{0} \tag{9-361}$$

不等式表明了左边的矩阵至少是 PSD 的：对任意非零向量 $\boldsymbol{u}\in\mathcal{R}^{M}$：

$$\boldsymbol{u}^{\mathrm{T}}(\operatorname{cov}[\boldsymbol{\theta}_{\mathrm{LS}}]-\operatorname{cov}[\hat{\boldsymbol{\theta}}])\boldsymbol{u}\geqslant 0 \tag{9-362}$$

为了得到$\hat{\boldsymbol{\theta}}$的协方差矩阵，注意到由式(9-360)：

$$\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}+\boldsymbol{B}\boldsymbol{V} \tag{9-363}$$

因此

$$\operatorname{cov}[\hat{\boldsymbol{\theta}}]=\mathcal{E}[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^{\mathrm{T}}]=\boldsymbol{B}\,\mathcal{E}[\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}]\boldsymbol{B}^{\mathrm{T}}=\sigma_V^2\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} \tag{9-364}$$

从每个估计器中减掉均值，得到以下误差向量：

$$\boldsymbol{E}=[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{\theta}]-[\boldsymbol{B}\boldsymbol{Y}-\boldsymbol{\theta}]=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}-\boldsymbol{B}\boldsymbol{V} \tag{9-365}$$

其中 $\boldsymbol{\theta}$ 项被消去了，因为所有的估计器都是无偏的。误差向量的协方差矩阵 $\boldsymbol{R}_{EE}=\mathcal{E}[\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}]$为

$$\begin{aligned}\boldsymbol{R}_{EE}&=[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}-\boldsymbol{B}]\varepsilon[\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}][(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}-\boldsymbol{B}]^{\mathrm{T}}\\&=\sigma_V^2[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}-\boldsymbol{B}((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}]\\&=\sigma_V^2[\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}]\end{aligned} \tag{9-366}$$

其中第二行等式中的前三项合成了一项，因为 $\boldsymbol{B}\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{I}$，$\boldsymbol{B}$ 是无偏的。由于 $\boldsymbol{R}_{EE}$ 是协方差矩阵，它至少是 PSD，则

$$\boldsymbol{R}_{EE}=\operatorname{cov}[\hat{\boldsymbol{\theta}}]-\operatorname{cov}[\boldsymbol{\theta}_{\mathrm{LS}}]\geqslant\mathbf{0} \tag{9-367}$$

这证明了 LS 估计器具有最小协方差：当 $\boldsymbol{R}_{VV}=\sigma_V^2\boldsymbol{I}$ 时，它是 BLU 估计器。如果噪声向量 $\boldsymbol{V}$ 也是高斯的，LS 估计器是所有线性和非线性估计器中最优的无偏估计器。

9.22　最优线性无偏估计

在 9.21 节中，我们假设 $\boldsymbol{R}_{VV}=\sigma_V^2\boldsymbol{I}$，因此 LS 估计器是 BLU 估计器。当噪声的协方差矩阵不具有对角阵形式，它仍可用加权矩阵 $\boldsymbol{W}=\boldsymbol{R}_{VV}^{-1}$ 构造的 WLS 来得到 BLU 估计。最终的估计器是：

$$\boldsymbol{\theta}_{\mathrm{WLS}}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{Y}=\boldsymbol{\theta}_{\mathrm{BLU}} \tag{9-368}$$

这个结果可在以下条件中采用一个约束的最优化推导得到。

- $\boldsymbol{Y}$ 的观测模型由式(9-310)给出，其中 $\boldsymbol{\theta}\in\mathcal{R}^{M}$ 是非随机的，但为未知量，$\boldsymbol{A}\in\mathcal{R}^{M\times N}$ 是已知的(可测的)，$\boldsymbol{V}\in\mathcal{R}^{N}$ 是一个随机向量，均值为零，协方差矩阵 $\boldsymbol{R}_{VV}\in\mathcal{R}^{N\times N}$
- BLU 估计器的形式是测量值的线性组合：

$$\boldsymbol{\theta}_{\mathrm{BLU}}=\boldsymbol{B}\boldsymbol{Y} \tag{9-369}$$

　其中 $\boldsymbol{B}\in\mathcal{R}^{M\times N}$。
- 它是无偏的：

$$\mathcal{E}[\boldsymbol{\theta}_{\text{BLU}}] = \mathcal{E}[\boldsymbol{B}(\boldsymbol{A\theta}+\boldsymbol{V})] = \boldsymbol{BA\theta}+\mathcal{E}[\boldsymbol{V}] = \boldsymbol{BA\theta} \tag{9-370}$$

需要 $\boldsymbol{BA}=\boldsymbol{I}$。

- 它具有最小协方差 $\operatorname{cov}[\boldsymbol{\theta}_{\text{BLU}}]$。目标是解出以下的约束最小值：

$$\boldsymbol{\theta}_{\text{BLU}} = \arg\min_{\boldsymbol{B}} \operatorname{cov}[\boldsymbol{\theta}_{\text{BLU}}] \text{ s. t. } \boldsymbol{BA} = \boldsymbol{I} \tag{9-371}$$

用这些假设，我们继续推导 BLU 估计器。由于它是无偏的，我们可以写做：

$$\boldsymbol{\theta}_{\text{BLU}} = \boldsymbol{\theta}+\boldsymbol{BV} \tag{9-372}$$

因此

$$\operatorname{cov}[\boldsymbol{\theta}_{\text{BLU}}] = \mathcal{E}[(\boldsymbol{\theta}_{\text{BLU}}-\boldsymbol{\theta})(\boldsymbol{\theta}_{\text{BLU}}-\boldsymbol{\theta})^{\mathrm{T}}] = \boldsymbol{B}\,\mathcal{E}[\boldsymbol{VV}^{\mathrm{T}}]\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{BR}_{\boldsymbol{VV}}\boldsymbol{B}^{\mathrm{T}} \tag{9-373}$$

将以上形式的协方差矩阵最小化等同于使其迹最小化：

$$\min_{\boldsymbol{B}} \operatorname{tr}[\boldsymbol{BR}_{\boldsymbol{VV}}\boldsymbol{B}^{\mathrm{T}}] \text{ s. t. } \boldsymbol{BA} = \boldsymbol{I} \tag{9-374}$$

则 BLU 估计器的代价函数为：

$$\xi_{\text{BLU}} \triangleq \operatorname{tr}[\boldsymbol{BR}_{\boldsymbol{VV}}\boldsymbol{B}^{\mathrm{T}}] \tag{9-375}$$

定义 $\boldsymbol{b}_m^{\mathrm{T}}\in\mathcal{R}^M$ 为 $\boldsymbol{B}$ 的第 m 行，迹中的每个元素按下式被最小化：

$$\min_{\boldsymbol{b}_m}(\boldsymbol{b}_m^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}\boldsymbol{b}_m+[\boldsymbol{b}_m^{\mathrm{T}}\boldsymbol{A}-\boldsymbol{e}_m^{\mathrm{T}}])\boldsymbol{\lambda}_m) \tag{9-376}$$

其中 $\boldsymbol{e}_m\in\mathcal{R}^M$ 为单位向量，第 m 个位置为 1，$\boldsymbol{\lambda}_m\in\mathcal{R}^M$ 是一个拉格朗日乘子向量。

用附录 G 中的微分法则：

$$\frac{\partial}{\partial \boldsymbol{b}_m}(\boldsymbol{b}_m^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}\boldsymbol{b}_m+[\boldsymbol{b}_m^{\mathrm{T}}\boldsymbol{A}-\boldsymbol{e}_m^{\mathrm{T}}]\boldsymbol{\lambda}_m) = 2\boldsymbol{R}_{\boldsymbol{VV}}\boldsymbol{b}_m+\boldsymbol{A\lambda}_m \tag{9-377}$$

将其置零，解出：

$$\boldsymbol{b}_m = -(1/2)\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A\lambda}_m \tag{9-378}$$

将这个结果代入约束条件，得到一个拉格朗日乘子的表达式：

$$-(1/2)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A\lambda}_m = \boldsymbol{e}_m \Rightarrow \boldsymbol{\lambda}_m = -2(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A})^{-1}\boldsymbol{e}_m \tag{9-379}$$

因此式(9-378)变为：

$$\boldsymbol{b}_m = \boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A})^{-1}\boldsymbol{e}_m \tag{9-380}$$

这个表达式正好满足约束条件：

$$\boldsymbol{b}_m^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{e}_m^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A} = \boldsymbol{e}_m^{\mathrm{T}} \tag{9-381}$$

由于 $\boldsymbol{b}_m^{\mathrm{T}}$ 是 $\boldsymbol{B}_{\text{BLU}}$ 的一行，$\boldsymbol{b}_m$ 是 $\boldsymbol{B}_{\text{BLU}}^{\mathrm{T}}$ 的一列。我们得到

$$\boldsymbol{B}^{\mathrm{T}} = [\boldsymbol{b}_1,\cdots,\boldsymbol{b}_M] = \boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A})^{-1}[\boldsymbol{e}_1,\cdots,\boldsymbol{e}_M] \tag{9-382}$$

其中 $[\boldsymbol{e}_1,\cdots\boldsymbol{e}_M]=\boldsymbol{I}\in\mathcal{R}^{M\times M}$ 是一个单位矩阵。因此

$$\boldsymbol{B} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{VV}}^{-1} \tag{9-383}$$

则得到式(9-368)中的 BLU 估计器。估计的样本用 $\hat{\boldsymbol{Y}}=\boldsymbol{A\theta}_{\text{BLU}}$ 产生。

注意到它与 LS 和 WLS 估计器不同，BLU 估计器是基于 $\boldsymbol{Y}$ 的基础模型，加性噪声的协方差阵 $\boldsymbol{R}_{\boldsymbol{VV}}$ 假设是已知的。$\boldsymbol{\theta}_{\text{BLU}}$ 的代价函数 ξ_{BLU} 不直接包括 $\boldsymbol{Y}$。LS 和 WLS 的代价函数的最小化是对参数向量 $\boldsymbol{\theta}$ 进行的操作，但这里是对 $\boldsymbol{\theta}_{\text{BLU}}$ 的行进行操作。样本只在式(9-369)中 $\boldsymbol{\theta}_{\text{BLU}}$ 的估计中出现。因此 BLU 估计器在实际中通常不是很有用。LS 和 WLS 估计器应用的更广泛，因为它们不需要关于 $\boldsymbol{Y}$ 的基础模型。以前用到的模型只是用来推导 LS 和 WLS 估计器的某些特性，假设它是样本在实际中如何产生的一个有效的代表。可能 BLU 估计器最重要的特点是 LS 和 WLS 估计器可得到其性能：正如第 9.23 节中总结的，它们在某些条件下是相同的估计器。

例 9-47 $M=2$ 时 $\boldsymbol{B}$ 的第一行的代价函数 ξ_{BLU}，与 $\boldsymbol{b}_1=[b_{11},\ b_{12}]^{\mathrm{T}}$ 对应，对小的样本数 $N=2$ 的一个例子在图 9-20a 中显示。噪声的协方差阵为：

$$\boldsymbol{R}_{\boldsymbol{VV}} = \begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix} \tag{9-384}$$

观察到平面是二次型的，约束是平面的，与碗形的平面相交。有效的 BLU 估计器的

解出现在约束平面与 $b_{11}-b_{12}$ 平面在 $\xi_{\text{BLU},1}=0$ 的相交处，以满足约束。二次函数的最优值为平面的最小值，位于两个平面相交的上面，用箭头标记了。在图 9-20b 中，约束平面被最优权值评估的代价函数抬高，显示它们在 $\boldsymbol{\theta}_{\text{BLU}}$ 处相交。这种图形对包括一个二次型的代价函数和参数服从线性约束的约束优化问题是非常典型的。

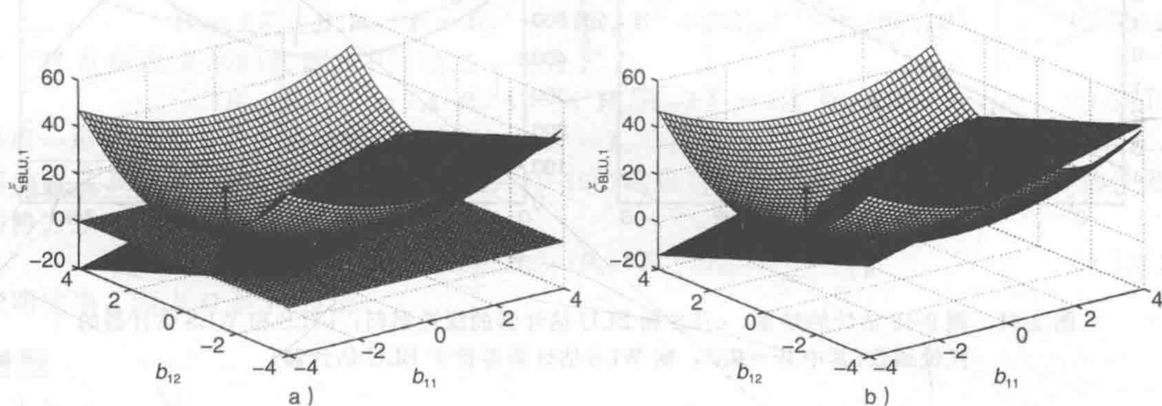


图 9-20　$N=2$ 时 BLU 估计器的代价函数和约束平面。a) 代价函数的真实位置和与 $b_{11}-b_{12}$ 平面在零水平面相关的约束平面；b) 为了显示两个函数在 $\hat{\boldsymbol{\theta}}_{\text{BLU}}$（由箭头给定）处相交，约束平面被最优权值处的代价函数值抬高，去掉了零水平面的 $b_{11}-b_{12}$ 平面　◀

例 9-48　再次考虑式(9-318)中的单参数模型($M=1$)，假设加性噪声向量具有非零元素 $\{\sigma_{V,n}^2\}$ 的对角线协方差阵。我们将证明 LS 和 BLU 估计器在噪声的方差随样本改变时的不同。对这种情况，式(9-383)中可逆项为：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A}=\sum_{n=1}^{N}X_n^2/\sigma_{V,n}^2 \tag{9-385}$$

因此

$$\boldsymbol{B}=\frac{1}{\sum_{n=1}^{N}X_n^2/\sigma_{V,n}^2}[X_1/\sigma_{V,1}^2,\cdots,X_N/\sigma_{V,N}^2]=\boldsymbol{b}^{\mathrm{T}} \tag{9-386}$$

它是对应单一参数的一个行向量。给出如下的 BLU 估计器：

$$\theta_{\text{BLU}}=\frac{\sum_{n=1}^{N}X_nY_n/\sigma_{V,n}^2}{\sum_{n=1}^{N}X_n^2/\sigma_{V,n}^2} \tag{9-387}$$

当 $\{\sigma_{V,n}^2\}$ 为相同值时简化为式(9-320)中的 $\hat{\theta}_{\text{LS}}$。对一般情况，在 BLU 估计器中具有更大噪声方差的样本的权重更小。图 9-21a 中显示的是 $M=1$ 对应的 BLU 和 LS 估计器生成的示例线条，$N=100$ 样本的噪声方差在[0,10]上均匀分布。两个线条不相同，因为 BLU 估计器考虑了噪声方差随样本的改变。真实的参数值是 $\theta=2$，两个估计是 $\theta_{\text{LS}}=1.5657$ 和 $\theta_{\text{BLU}}=1.9362$。显然 BLU 估计器更精确。当然，正如之前所讨论的，想要实际中每个样本的噪声方差为已知是不现实的，尽管一些信息是可利用的，比如噪声方差是否是增长的。由于估计器是随机变量，对某个特定的样本集，LS 估计器可能比 BLU 估计器更精确。尽管 θ_{BLU} 的方差更低些：对这个例子，它是 0.0225，而 θ_{LS} 为 0.0564。图 9-21b显示的是 LS 代价函数，将它与权值矩阵 $\boldsymbol{W}$ 包含变化的噪声方差时的 WLS 代价函数比较(回顾一下 WLS 和 BLU 估计器在 $\boldsymbol{W}=\boldsymbol{R}_{VV}^{-1}$ 时是相同的)。与上面的估计器对应的曲线的最小值由垂直的虚线标示。尽管 WLS/BLU 曲线在 LS 曲线的下方，但这不是一定会发生的情况。这两个代价函数的相关值不一定都是有含义的；它们的最小值和对应的估计器是非常重要的。

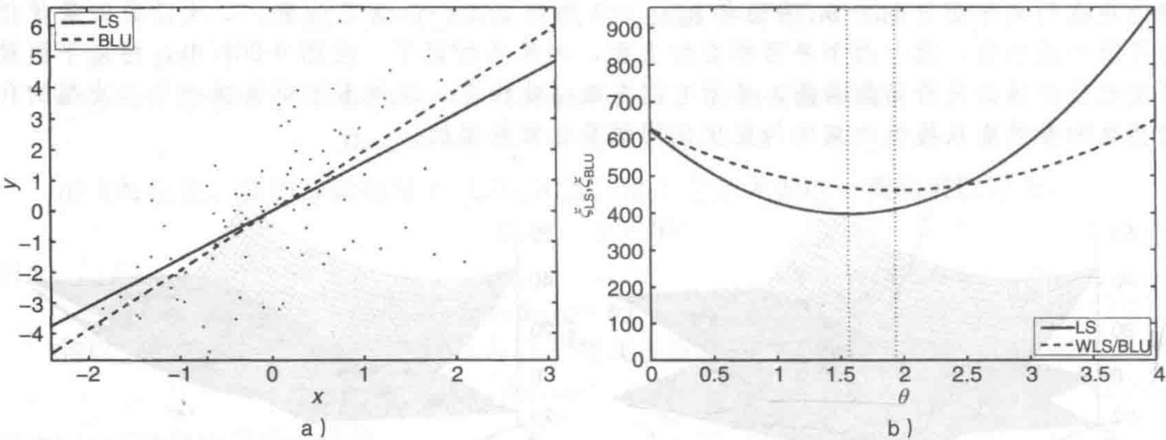


图 9-21 例 9-48 估计的结果。a)LS 和 BLU 估计器的线性回归；b)LS 和 WLS 估计器的代价函数(其中 $\boldsymbol{W}=\boldsymbol{R}_{VV}^{-1}$，则 WLS 估计器等价于 BLU 估计器) ◀

9.23 BLU 估计器的性质

尽管 BLU 估计器是用一种不同于 WLS 最小化的方法推导出来的，但实际上结果只是这种方法的一种特殊情况

- BLU 和 WLS 估计器的等价性。当 $\boldsymbol{W}=\boldsymbol{R}_{VV}^{-1}$ 时，BLU 和 WLS 估计器是相同的：

$$\boldsymbol{\theta}_{\mathrm{WLS}}\big|_{W=R_{VV}^{-1}}=\boldsymbol{\theta}_{\mathrm{BLU}} \tag{9-388}$$

  当 $\boldsymbol{V}$ 是不相关的，即 $\boldsymbol{R}_{VV}=\sigma_V^2\boldsymbol{I}$，LS 和 BLU 估计器是等价的。这是之前给出的 LS 估计器的特性的重述。

- 无偏性。由于 BLU 估计器的推导包括了无偏性的约束，$\boldsymbol{\theta}_{\mathrm{BLU}}$ 对零均值的噪声显然是无偏的：

$$\begin{aligned}\mathcal{E}[\boldsymbol{\theta}_{\mathrm{BLU}}]&=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\,\mathcal{E}[\boldsymbol{Y}]\\&=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}(\boldsymbol{A}\boldsymbol{\theta}+\mathcal{E}[\boldsymbol{V}])=\boldsymbol{\theta}\end{aligned} \tag{9-389}$$

- 协方差矩阵。BLU 估计器的协方差矩阵为

$$\begin{aligned}\mathrm{cov}[\boldsymbol{\theta}_{\mathrm{BLU}}]&=\mathcal{E}[(\boldsymbol{\theta}_{\mathrm{BLU}}-\boldsymbol{\theta})(\boldsymbol{\theta}_{\mathrm{BLU}}-\boldsymbol{\theta})^{\mathrm{T}}]\\&=\mathcal{E}[\boldsymbol{B}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}]=\boldsymbol{B}\boldsymbol{R}_{VV}\boldsymbol{B}^{\mathrm{T}}\end{aligned} \tag{9-390}$$

代入式(9-368)得到：

$$\begin{aligned}\mathrm{cov}[\boldsymbol{\theta}_{\mathrm{BLU}}]&=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{R}_{VV}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1}\\&=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1}\end{aligned} \tag{9-391}$$

- 最小 ξ_{BLU}。代价函数的最小值是式(9-391)中的协方差阵的迹：

$$\xi_{\mathrm{BLU,min}}=\mathrm{tr}[\boldsymbol{B}\boldsymbol{R}_{VV}\boldsymbol{B}^{\mathrm{T}}]=\mathrm{tr}[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1}] \tag{9-392}$$

  有意思的是通常 WLS 估计器代价函数的最小值依赖于样本 $\boldsymbol{Y}$，而此处不是这种情况。这个结果是源于 BLU 估计器是通过将基于 $\boldsymbol{Y}$ 的一个基础模型的协方差阵最小化推导出来的。

- 有效性。最后，我们证明 BLU 估计器是最有效的线性无偏估计器，它是“最好的”，正如其名称中包含的意思。

定理 9-9 BLU 估计器 $\boldsymbol{\theta}_{\mathrm{BLU}}$ 在所有线性无偏估计器中具有最小方差。

证明： 定义以下差矩阵：

$$\boldsymbol{D}\triangleq\mathrm{cov}[\hat{\boldsymbol{\theta}}]-\mathrm{cov}[\boldsymbol{\theta}_{\mathrm{BLU}}] \tag{9-393}$$

其中 $\hat{\boldsymbol{\theta}}$ 是另一个线性无偏估计器：它具有 $\boldsymbol{FY}$ 的形式，$\mathcal{E}[\boldsymbol{FY}]=\boldsymbol{\theta}$。目的是证明 $\boldsymbol{D}$ 是每个非零向量 $\boldsymbol{u}$ 的 PSD：$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{u}\geqslant0$，在这种意义下，BLU 估计器的协方差是最小的。用类似的方法证明 LS 估计器在 $\boldsymbol{R}_{VV}=\boldsymbol{I}$ 时是 BLU 估计器。此处的结果可用于任意噪声协方差阵

$\boldsymbol{R}_{VV}$。这两个估计器的协方差矩阵具有相同的形式：

$$\mathrm{cov}[\boldsymbol{\theta}_{\mathrm{BLU}}]=\boldsymbol{BR}_{VV}\boldsymbol{B}^{\mathrm{T}},\quad \mathrm{cov}[\hat{\boldsymbol{\theta}}]=\boldsymbol{FR}_{VV}\boldsymbol{F}^{\mathrm{T}} \tag{9-394}$$

当代入 $\boldsymbol{B}_{\mathrm{BLU}}$，我们需要确保两个估计器都是无偏的，这意味着 $\boldsymbol{BA}=\boldsymbol{FA}=\boldsymbol{I}$。插入变量，重新求结果得到：

$$\boldsymbol{D}=(\boldsymbol{F}-\boldsymbol{B})\boldsymbol{R}_{VV}(\boldsymbol{F}-\boldsymbol{B})^{\mathrm{T}}+(\boldsymbol{FR}_{VV}\boldsymbol{B}^{\mathrm{T}}+\boldsymbol{BR}_{VV}\boldsymbol{F}^{\mathrm{T}})-2\boldsymbol{BR}_{VV}\boldsymbol{B}^{\mathrm{T}} \tag{9-395}$$

将 B 用式(9-368)代替，中间项之一变为了

$$\boldsymbol{B}_{\mathrm{BLU}}\boldsymbol{R}_{VV}\boldsymbol{F}^{\mathrm{T}}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{R}_{VV}\boldsymbol{F}^{\mathrm{T}}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{VV}^{-1}\boldsymbol{A})^{-1} \tag{9-396}$$

最后一项用到了无偏性的约束 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}=(\boldsymbol{FA})^{\mathrm{T}}=\boldsymbol{I}$。对其他的中间项也可得到同样的结果，因为它是 $\boldsymbol{BR}_{VV}\boldsymbol{F}^{\mathrm{T}}$ 的转置。将 $\boldsymbol{B}_{\mathrm{BLU}}$ 代入式(9-395)的最后一项得到式(9-391)中 BLU 估计器的协方差，因此四项消掉了，得到

$$\boldsymbol{D}=(\boldsymbol{F}-\boldsymbol{B}_{\mathrm{BLU}})\boldsymbol{R}_{VV}(\boldsymbol{F}-\boldsymbol{B}_{\mathrm{BLU}})^{\mathrm{T}} \tag{9-397}$$

证明完毕，因为 $\boldsymbol{D}$ 是 PSD 的。

习题

充分统计量和 UMVU 估计

9-1 对 N 个 iid 泊松样本，用充分性的定义找参数 α 的一个充分统计量。

9-2 对 N 个 iid 均匀样本，pdf 为 $f_X(x)=I_{[\theta_1,x]}(x)I_{[x,\theta_2]}(x)$，用因式定理找$\{\theta_1,\theta_2\}$的一个充分统计量。

9-3 重复习题 9-2 的问题，对 N 个 iid 伽马样本找到 $\boldsymbol{\theta}=\{\alpha, r\}$的一个充分统计量。

9-4 推导式(9-30)中的期望值来求最大和最小值。

9-5 求 M 个参数为 N 的具有卡方分布的 iid 样本的最小充分统计量。

9-6 用式(9-34)中比值方法从 N 个参数为 α 的 iid 泊松随机变量样本中找最小充分统计量。

9-7 对位置参数为 c 的指数随机变量：

$$f_X(x)=\lambda\exp(-\lambda(x-c))I_{[c,\infty)}(x) \tag{9-398}$$

证明 c(固定 λ)的最小充分统计量是 N 个 iid 样本中的 $X_{(1)}$。

9-8 对 N 个参数为 $\{\mu,\alpha\}$ 的 iid 拉普拉斯样本，求以下参数的最小充分统计量：(a)α(固定 $\mu=0$)；(b)μ(固定 $\alpha=1$)。

9-9 对 N 个在 $[0,\theta]$ 上 iid 的均匀样本，$X_{(N)}$ 是 θ 的最小充分统计量。当 $1\leqslant m\leqslant n\leqslant N$ 时，$\{X_{(m)},X_{(n)}\}$ 的联合 pdf 在习题 4-41 的式(4-269)中给定。计算拉奥-布莱克维尔定理中的$\mathcal{E}[X_{(m)}\mid X_{(N)}]$，指定函数 $c(\theta)$。

9-10 证明样本均值 $\overline{X}$ 是已知 σ 的高斯分布中 μ 的完备充分统计量。

9-11 对 N 个参数为 $\{\mu,\sigma=\mu\}$ 的 iid 高斯样本，找 μ 的最小充分统计量。

9-12 对 N 个参数为 α 的半正态 iid 样本，求 α^2 的 UMVU 估计。

9-13 对 M 个参数为 $\{N,p\}$ 的二项 iid 样本，求 p^2 的 UMVU 估计(假设 $N>2$ 已知)。

9-14 位置 θ 的皮特曼估计可由 iid 样本得到，如下式所示：

$$T=\frac{\int_{-\infty}^{\infty}\theta\prod_{n=1}^{N}f_X(x_n-\theta)\mathrm{d}\theta}{\int_{-\infty}^{\infty}\prod_{n=1}^{N}f_X(x_n-\theta)\mathrm{d}\theta} \tag{9-399}$$

求 X 为参数 $\{\mu,\sigma\}$ 的高斯(σ 已知)时 $\theta=\mu$ 的皮特曼估计。

MSE 估计

9-15 对加性噪声问题 $X=Y+V$，其中 Y 和 V 是独立随机变量，求当 V 是参数 $\{\mu=0,\sigma_V\}$ 的高斯分布，Y 是参数为 p 输出为$\{\pm a\}$的对称伯努利时，Y 的 MSE、MAE 和 ML 估计。

9-16 对习题 9-15，设 MSE 估计受以下仿射形式约束：

$$g_{\mathrm{MSE}}(X)=aX+b \tag{9-400}$$

假设 V 的均值为 $\mu=1$，求使 MSE 最小的 a 和 b。

9-17 重复习题 9-15，当噪声 V 具有参数为 $\{\mu=0,\alpha\}$ 的拉普拉斯分布。

9-18 设随机序列 $X[k](k\in\mathcal{N})$ 的自相关函数为 $R_{XX}[m]=(2/3)^{|m|}$。考虑以下 MSE 估计：

$$X_{\mathrm{MSE}}[k]=aX[1]+bX[N] \tag{9-401}$$

其中 $k\in\{2,\cdots,N-1\}$。求使$\mathcal{E}[(X[k]-X_{\mathrm{MS}}[k])^2]$最小的最优参数 $\{a,b\}$。

9-19 假设

$$X_{\mathrm{MSE}}=\sum_{m=1}^{M}a_mY_m=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{Y} \tag{9-402}$$

为一个基于 M 个相关随机变量$\{Y_m\}$的 X 的估计，其中 $\boldsymbol{a}\triangleq[a_1,\dots,a_M]^{\mathrm{T}}$，$\boldsymbol{Y}=[Y_1,\cdots,Y_M]^T$。求使 $\varepsilon[(X-X_{\mathrm{MSE}}[k])^2]$最小的 $\boldsymbol{a}$ 的表达式。

MAP 和 ML 估计

9-20 求参数为 α 的泊松分布的 MAP 估计，假设

α是一个标准高斯随机变量。

9-21 证明式(9-287)。

9-22 设测量值按以下方式产生：

$$Y = X + 2A + A^2 \quad (9\text{-}403)$$

其中 X 是参数 $\lambda=1$ 的指数分布，A 具有以下 pdf：

$$f_A(a) = (1/4)[3\delta(a-1) + \delta(a-3)] \quad (9\text{-}404)$$

对 Y 的单次测量，求：(a)A 的 ML 估计；(b)A的 MAP 估计。

9-23 求基于 X 的 N 个样本的 θ 的 ML 估计，X 是$[0,\theta]$上均匀分布的。

9-24 重复习题 9-23，X 具有参数为 $\{a=0,b=1,c\}$ 的三角分布。

9-25 由参数为 $\{\theta=r,\alpha=1\}$ 的韦伯分布的 N 个样本推导 θ 的 ML 估计的表达式。

9-26 考虑随机变量 X 分布的以下两个假设。H_0：参数为 $\{\mu=0,\alpha\}$ 的高斯分布。H_1：参数为 $\{\mu=0,\alpha\}$ 的拉普拉斯分布。用 LRT 来求 $\eta=1$ 时 x 的测试条件的表达式。在图 3-20中给出了 pdf 的对比。

9-27 重复习题 9-26，此时 H_1 为参数 $\{c=0,\alpha\}$ 的柯西分布。

9-28 推导估计一个具有已知方差的有限高斯混合 $\boldsymbol{\theta}=[a_{1,\dots},a_M,\mu_{1,\dots},\mu_M]^{\mathrm{T}}$ 的 EM 算法，与导出式(9-289)和式(9-295)的过程类似。

矩方法

9-29 求由参数为 $\{\mu,\alpha\}$ 的逆高斯分布的 N 个 iid 样本得出的 MoM 估计。

9-30 重复习题 9-29，针对参数为 $\{N,P\}$ 的负二项随机变量。

9-31 重复习题 9-29，针对参数为 $\{\mu=1,\alpha\}$ 的拉普拉斯随机变量。

克拉美罗下界(CRLB)

9-32 (a)求定义域为$\mathcal{Z}^+$几何分布的 p 的无偏估计的 CRLB；(b)推导 p 的 ML 估计。

9-33 求参数为$\{p, N\}$的二项随机变量中 p 的 Fisher 信息。

9-34 重复习题 9-33，参数为 $\{\lambda,r\}$ 的伽马分布中的 λ。

9-35 推导附录 F 式(F-71)中偏移 $B(\theta)$的有偏估计的 CRLB 的结果。

最小平方估计

9-36 LS 估计具有形式$\hat{\boldsymbol{Y}}_{\mathrm{WLS}}=\boldsymbol{PY}$，其中 $\boldsymbol{P}\triangleq\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{WA})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}$ 是一个投影矩阵，因为如图 9-18所示，它将 $\boldsymbol{Y}$ 投影在 $\boldsymbol{A}$ 的范围内。显然$\widetilde{\boldsymbol{Y}}_{\mathrm{WLS}}=\boldsymbol{Y}-\hat{\boldsymbol{Y}}_{\mathrm{WLS}}=(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{Y}$。证明 $\boldsymbol{P}$ 是一个幂等矩阵，它与 $\boldsymbol{I}-\boldsymbol{P}$ 正交。

9-37 设 LS 估计受以下条件限制：

$$\boldsymbol{\theta}_{\mathrm{LS}} = \min_{\theta}[\widetilde{\boldsymbol{Y}}^{\mathrm{T}}\widetilde{\boldsymbol{Y}} + \lambda(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{\theta}-c)] \quad (9\text{-}405)$$

其中 λ 为一个拉格朗日乘子，约束为对 $\boldsymbol{c}\in\mathcal{R}^M$，满足 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{\theta}=c$。求 $\boldsymbol{\theta}_{\mathrm{LS}}$的表达式。

9-38 假设 LS 模型扩展为：

$$\boldsymbol{Y} = \boldsymbol{A\theta} + \boldsymbol{b} + \boldsymbol{V} \quad (9\text{-}406)$$

其中 $\boldsymbol{b}\in\mathcal{R}^N$ 为一个固定的已知向量。求此模型的加权 LS 的公式，求 $\boldsymbol{\theta}_{\mathrm{LS}}$的表达式。

仿真作业

9-39 用 MATLAB 完成此章描述的对有穷混合模型的 EM 算法，分别用式(9-291)和式(9-295)给出的第 E 步和第 M 步。假设

$$f_X(x) = (0.3)(1/2)\exp(-x/2)u(x) + (0.7)\sqrt{2/\pi}\exp(-x^2/2)u(x) \quad (9\text{-}407)$$

是指数($\lambda=1/2$)和半正态($\alpha=1$)分布的混合。画出实际的和估计的混合分布图。

9-40 考虑以下模型：

$$Y = \theta_1 + \theta_2 X - \theta_3 X^3 + V \quad (9\text{-}408)$$

其中 X 是$[0, 1]$上均匀分布的，V 是参数为 $\{\mu=0,\sigma=0.1\}$ 的高斯分布。尽管 X^3 出现在这个表达式中，它仍是一个关于参数的线性模型。求对三个参数 $\{\theta_1=1,\theta_2=4,\theta_3=-2\}$ 具有单位加权矩阵 $\boldsymbol{W}=\boldsymbol{I}$ 的 WLS 估计。用 MATLAB 来画 $\{X,Y\}$ 的 $N=500$ 个样本和估计 $\boldsymbol{\theta}_{\mathrm{LS}}$生成的线条，与图 9-16 中显示的类似。

进一步阅读

本章中涵盖的问题在以下参考书中有进一步深入讨论。充分统计量和估计：Haykin (2002)，Lehman (1983)，Kailath、Sayed 和 Hassibi (2000)，Kay (1993)，Mendel (1995)，Poor (1994)，Sorenson (1980)，Van Trees (1968)。主要分量回归：Jolliffe (1986)，Mosteller 和 Tukey (1977)。

本书第三部分的注意事项

第三部分包括四章内容，介绍在信号处理和通信中的应用。它们可在 www.wiley.com/go/randomprocesses 网站中找到。以下是内容梗概：

第 10 章：通信系统和信息原理；

第 11 章：最优滤波；

第 12 章：自适应滤波；

第 13 章：均衡、波束形成和导航。

附录

附录内容介绍

在附录中，我们提供了一些对前面章节部分问题进行综述的背景资料，读者会发现它们非常有用。当然，其中大多数信息可以在本科工程课程使用的其他教科书(如频域变换的概述)中找到，但本书包括这些内容后会使得本书相对独立。

- 附录A　将第3章中描述的所有的单变量随机变量进行了总结以便于参考。提供了概率密度函数(pdf)、累积分布函数(cdf)、矩、特征函数、熵的表达式以及每种分布的示例曲线(显示了均值和宽度 $\mu_X \pm \sigma_X$)。随机变量是按字母顺序排列的，且分开描述连续和离散分布。
- 附录B　涵盖了函数的连续性、有界性以及各种其他特性。包括了贯穿全书的特殊函数，并提供了示例图来说明它们的特性。几乎所有的这些函数都是第3章中涉及非常广泛的分布，以及需要以简洁的形式描述 pdf、cdf、矩和熵而引出的。
- 附录C　对以下频域变换的性质和变换对进行了概括：(ⅰ)拉普拉斯变换；(ⅱ)连续时间傅里叶变换；(ⅲ)z 变换；(ⅳ)离散时间傅里叶变换。讨论了双边和单边变换，还包括一个用于求 z 逆变换的柯西留数定理的例子。
- 附录D　包含了对黎曼积分的简要回顾，介绍了对于计算期望值非常有用的黎曼-斯蒂尔切斯和勒贝格积分。总结了连续随机变量的 pdf 的函数形式，这在计算一些积分时非常有用。展示了如何对一些两变量的“特殊”函数积分，比如 $\delta(x-y)$ 和 $u(x-y)$。
- 附录E　给出了各种恒等式，包括三角恒等式、斯特林公式以及一些级数展开和闭合形式的和。包括了与附录 D 中的 pdf 积分类似的离散随机变量的概率质量函数的和。
- 附录F　给出了几个不等式，它们对于计算边缘概率和期望非常有用。这些不等式包括柯西-施瓦茨和三角不等式、切而诺夫边界、杰森不等式以及克拉美罗下限，可用于第9章的参数估计中。
- 附录G　总结了矩阵和向量的重要性质。描述了一些矩阵分解、四个基本子空间和矩阵求逆定理。还通过例题给出了向量(实数和复数)微分的法则。

附录 A

单变量参数分布总结

A.1 符号

以下符号用于下文中对具有一个、两个或三个参数的随机变量分布的总结：

$f_X(x)$ 概率密度函数(pdf)

$p_X[x]$ 概率质量函数(pmf)

$F_X[x]$ 累积分布函数(cdf)

m_e 中位数

m_o 众数

μ_n n 阶原点(非中心)矩$\mathcal{E}[X^n]$

$\mu_{c,n}$ n 阶中心矩$\mathcal{E}[(X-\mu_X)^n]$

μ_X 均值$\mathcal{E}[X]=\mu_1$

σ_X^2 方差$\mathcal{E}[(X-\mu_X)^2]=\mu_{c,2}$

γ_1 偏度系数 $\mu_{c,3}/\sigma_X^3$

γ_2 超峰态系数 $\mu_{c,4}/\sigma_X^4-3$

$\Phi_X(\omega)$ 特征函数$\mathcal{E}[\exp(\mathrm{j}\omega X)]$(连续型随机变量)

$\Phi_X(\mathrm{j}\omega)$ 特征函数$\mathcal{E}[\exp(\mathrm{j}\omega X)]$(离散型随机变量)

$H(X)$ 熵(离散型随机变量)

$h(X)$ 微分熵(连续型随机变量)

对 $H(X)$和 $h(X)$，到底是用 $\log(x)$(任意底)还是 $\ln(x)$(以 e 为底)，取决于包含指数分布的熵的表达是否更简洁。对于熵的表达式的一般做法如下：(Ⅰ)对于 $\log(x)$，表达式中可用任何基底；(ⅱ)对 $\ln(x)$，表达式仅适用于自然对数，如果需要其他底，则必须要进行转换。

由于很难写出各种随机变量的所有矩(或中心矩)的通用表达式，所以只给出低阶的结果。单位阶跃函数 $u(x)$用来定义连续随机变量的定义域，指示函数用于第 3 章(对于三角形分布，指示函数使表达式更简洁。)对于离散随机变量、概率密度函数(pdf)和累积分布函数(cdf)分别写为狄拉克 δ 函数 $\delta(x)$和单位阶跃函数 $u(x)$的加权和。其结果是，一个离散随机变量只需用 $u(x)$替代 $\delta(x)$就可从其 pdf 得到其 cdf。为了使本文更完备，还给出了概率质量函数(pmf)，因为有些读者可能会更愿意用这个符号。

在附录 B 中描述的以下的函数和数在本附录中也用到了：

$B(a,b)$ 贝塔函数$=(\Gamma(a)\Gamma(b)/\Gamma(a+b)$

$B_y(a,b)$ 不完全贝塔函数

$\Gamma(a)$ 伽马函数

$\gamma(a,b)$ 下不完全伽马函数

$\Gamma(a,b)$ 上不完全伽马函数

$\psi(x)$ 双 Γ 函数

$\zeta(x)$ 泽塔函数

$I_{(a,b]}(x)$ 指示函数(连续 x)

$I_{\{a,b\}}[x]$ 指示函数(离散 x)

$\mathrm{sgn}(x)$ 符号函数

$Q(x)$ Q-函数

$\mathrm{erf}(x)$ 误差函数

$\mathrm{erfi}(x)$ 虚误差函数

$I_\alpha(x)$ 修正贝塞尔函数(第一类)

$K_\alpha(x)$ 修正贝塞尔函数(第二类)

$Q_m(a,b)$ 马库姆-Q 函数

$L_v(x)$ 拉盖尔多项式

${}_1F_1(a;\ b;\ x)$ 合流超几何函数(第一类)

$\Psi(a;\ b;\ x)$ 合流超几何函数(第二类)

$\lfloor x \rfloor$ 底函数

$\lceil r \rceil$ 顶函数

B_n 伯努利数

γ 欧拉-马歇罗尼常数(≈0.5772)

为了符号简单，本书中在除以一个乘积项时并不总是包括方括号或圆括号。例如，在指数项 $\exp(x^2/2\sigma^2)$ 中应该清楚是 x^2 除以 $2\sigma^2$，但在大多数计算机编程语言中操作顺序并不是这样的(包括 MATLAB)®。最后，总结中并不包括以下一些量，理由如下：

太复杂：表达式太复杂而且可能没有用(典型的有特征函数或熵)；

隐式：不能写出一个明确的表达式(如中位数或众数)；

不存在：矩是不确定的或无限的(柯西)(也是伯努利和离散均匀的中位数)；

不唯一：众数可以为许多值(均匀的)(二项分布和泊松分布可以取两个值)。

A.2 拓展阅读

没有一本参考资料包括了本附录中所有的信息。读者可以在本书结尾所列出的参考书目中找到相关内容。下面的书是非常有用的，尤其是关于矩和单变量分布的其他特性：Johnson，Kemp 和 Kotz(2005)；Johnson，Kotz 和 Balakrishnan(1994)；Johnson，Kotz 和 Balakrishnan(1995)。书中也总结了其他一些参考文献：Krishnamoorthy(2006)；Mood，Graybill 和 Boes(1974)；Pal，Jin 和 Lim(2006)；Peebles(1993)，Weisstein(2003)，Yates 和 Goodman(2005)；Zwillinger 和 Kokoska(2000)。

在这里没有对第 4 章涉及的多变量分布进行总结，读者可以参考以下资料：Gumbel(1960)；Johnson，Kotz 和 Balakrishnan(1997)；Kotz，Balakrishnan 和 Johnson(2000)；Mardia(1970)。

大多数微分熵的表达式可以在 Verdugo Lazo 和 Rathie(1978)中找到。

最后，我们提供了少量许多没有涉及概率的教科书中广泛讨论的特殊分布的资料。半正态：Leone，nelson 和 Nottingham(1961)。逆高斯：Chhikara 和 Folks(1989)；Seshadri(1999)。拉普拉斯：Kotz，Kozubowski 和 Podg′orski(2001)。逻辑：Balakrishnan 和 Ed(1992)。对数正态分布：Limpert，Stahel 和 Abbt(2001)。韦伯：Rousu(1973)。

A.3 连续随机变量

A.3.1 贝塔，$\alpha=\beta=1/2$ 的(Arcsine，$\beta=1$ 的功率函数)

参数 $\alpha,\beta>0$，定义域 $x\in[0,1]$

$$f_X(x)=[1/B(\alpha,\beta)]x^{\alpha-1}(1-x)^{\beta-1}[u(x)-u(x-1)]$$

$$F_X(x)=[B_x(\alpha,\beta)/B(\alpha,\beta)][u(x)-u(x-1)]+u(x-1)$$

$$\mu_X=\alpha/(\alpha+\beta),\quad \sigma_X^2=\alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2]$$

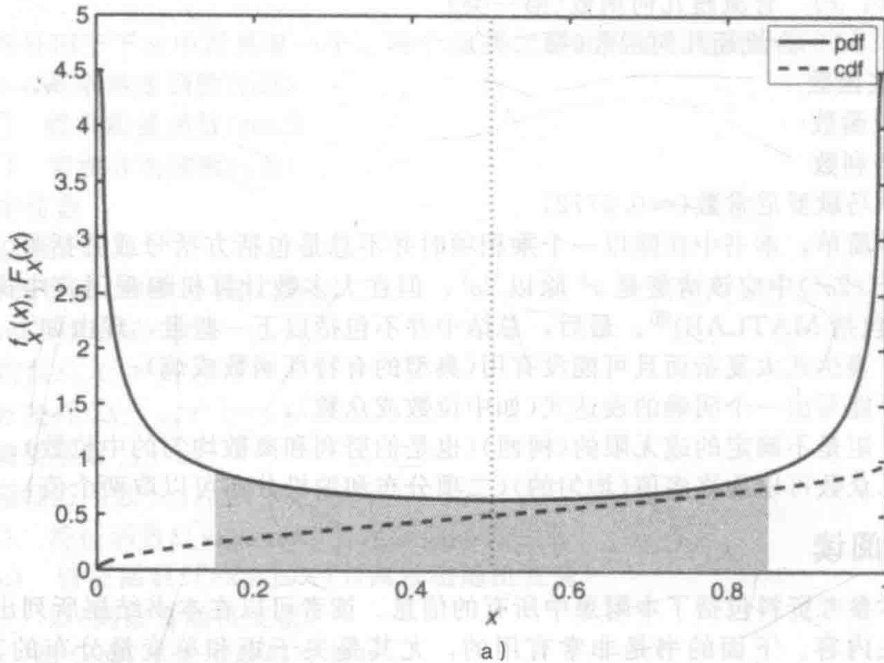
（续）

$$m_e \text{为隐式}, \quad m_o = (\alpha-1)/(\alpha+\beta-2)(\text{对于 } \alpha,\beta>1)$$

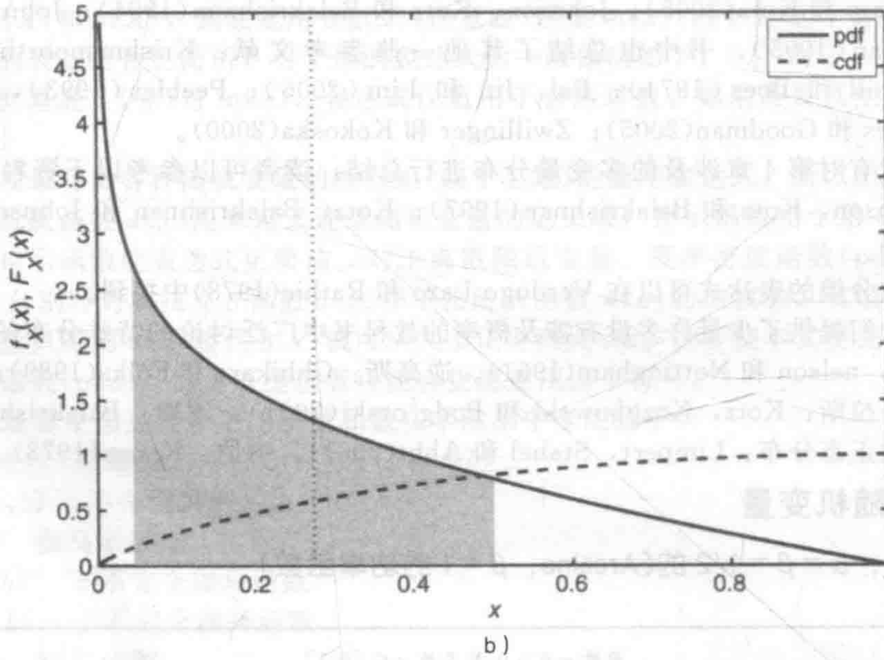
$$\mu_n = B(n+\alpha,\beta)/B(\alpha,\beta), \quad \gamma_1 = \sqrt{(\alpha+\beta+1)/\alpha\beta}\,\frac{2(\beta-\alpha)}{(\alpha+\beta+2)}$$

$$\gamma_2 = 6\,\frac{\alpha^3+\alpha^2(1-2\beta)-2\alpha\beta(\beta+2)+\beta^2(\beta+1)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}, \quad \Phi_X(\omega) = {}_1F_1(\alpha;\beta;j\omega)$$

$$h(X) = (\alpha+\beta-2)\psi(\alpha+\beta)-(\alpha-1)\psi(\alpha)-(\beta-1)\psi(\beta)+\ln(B(\alpha,\beta))$$



a）



b）

图 A-1　贝塔 pdf 和 cdf。a）贝塔随机变量 $\alpha=\beta=0.5$；b）贝塔随机变量 $\alpha=0.75$ 和 $\beta=2$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.3.2 柯西

$$\text{参数 } c \in \mathcal{R}, \alpha > 0, \text{定义域 } x \in \mathcal{R}$$
$$f_X(x) = (\alpha/\pi)[(x-c)^2 + \alpha^2]^{-1}$$
$$F_X(x) = 1/2 + (1/\pi)\arctan((x-c)/\alpha)$$
$$\mu_X = \text{不存在}, \quad \sigma_X^2 = \text{不存在}$$
$$m_e = c, \quad m_o = c$$
$$\mu_n \text{ 不存在}, \quad \mu_{c,n} \text{ 不存在},$$
$$\gamma_1 \text{ 不存在}, \quad \gamma_2 \text{ 不存在}$$
$$\Phi_X(\omega) = \exp(\mathrm{j}\omega c - \alpha|\omega|), \quad h(X) = \log(4\pi\alpha)$$

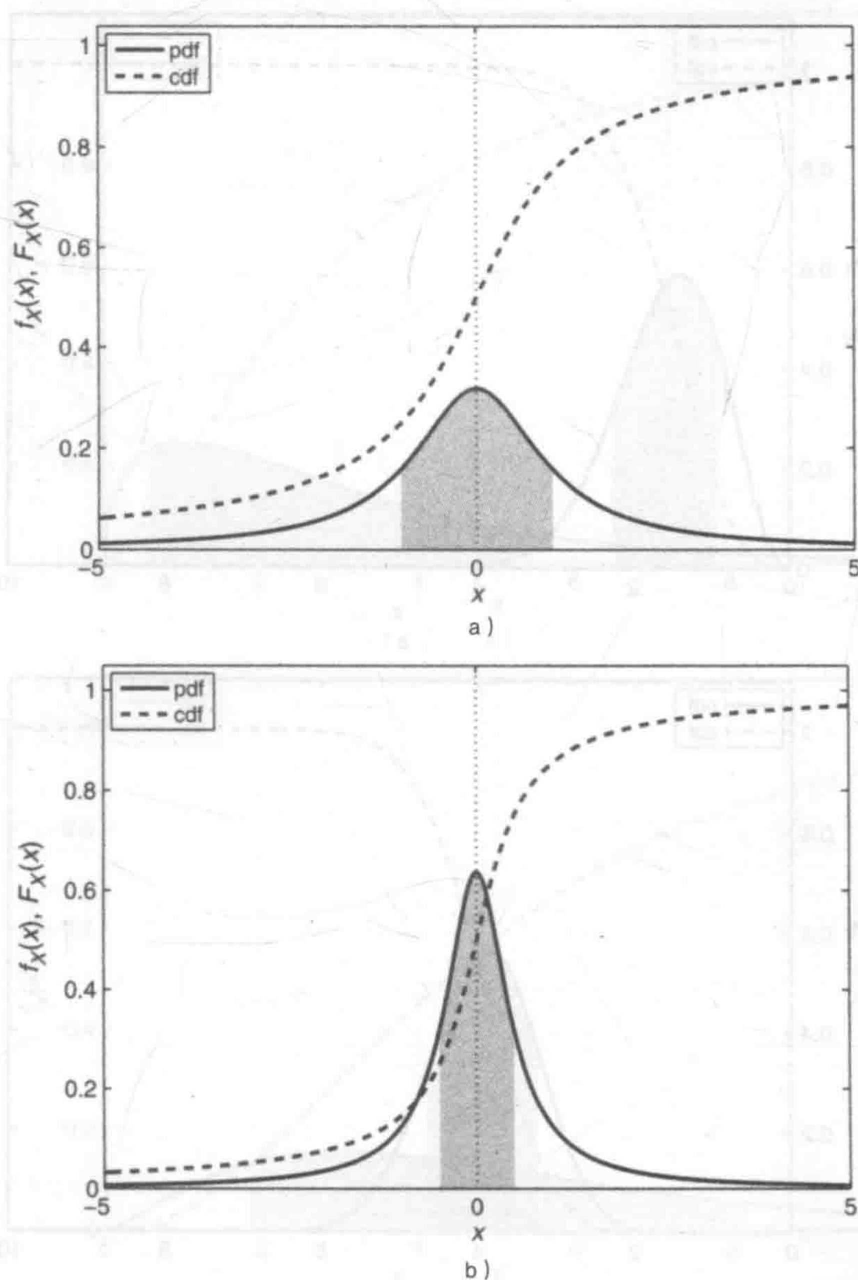


图 A-2 柯西 pdf 和 cdf。a) 柯西随机变量 $c=0$，$\alpha=1$；b) 柯西随机变量 $c=0$，$\alpha=0.5$。虚线表示 m_e，阴影部分表示 $[F_X^{-1}(1/4), F_X^{-1}(3/4)]$（由于 μ_X 和 σ_X 不存在）

A. 3. 3 Chi

参数 $N \in \mathcal{N}$，定义域 $x \in \mathcal{R}^+$

$$f_X(x) = [x^{N-1}/2^{N/2-1}\Gamma(N/2)]\exp(-x^2/2)u(x)$$

$$F_X(x) = [\gamma(N/2, x^2/2)/\Gamma(N/2)]u(x)$$

$$\mu_X = \sqrt{2}\Gamma((N+1)/2)/\Gamma(N/2), \quad \sigma_X^2 = N - 2[\Gamma((N+1)/2)/\Gamma(N/2)]^2$$

m_e 为隐式， $m_0 = \sqrt{N-1}$（其中 $N \geqslant 1$）

$\mu_n = 2^{n/2}\Gamma((N+n)/2)/\Gamma(N/2)$， γ_1 太复杂， γ_2 太复杂

$$\Phi_X(\omega) = [1/\Gamma(N/2)]\sum_{n=0}^{\infty}[(j\omega\sqrt{2})^n/n!)\Gamma((N+n)/2)$$

$$h(X) = (1/2)[(N-(N-1)\psi(N/2) - \ln(2)] + \ln(\Gamma(N/2))$$

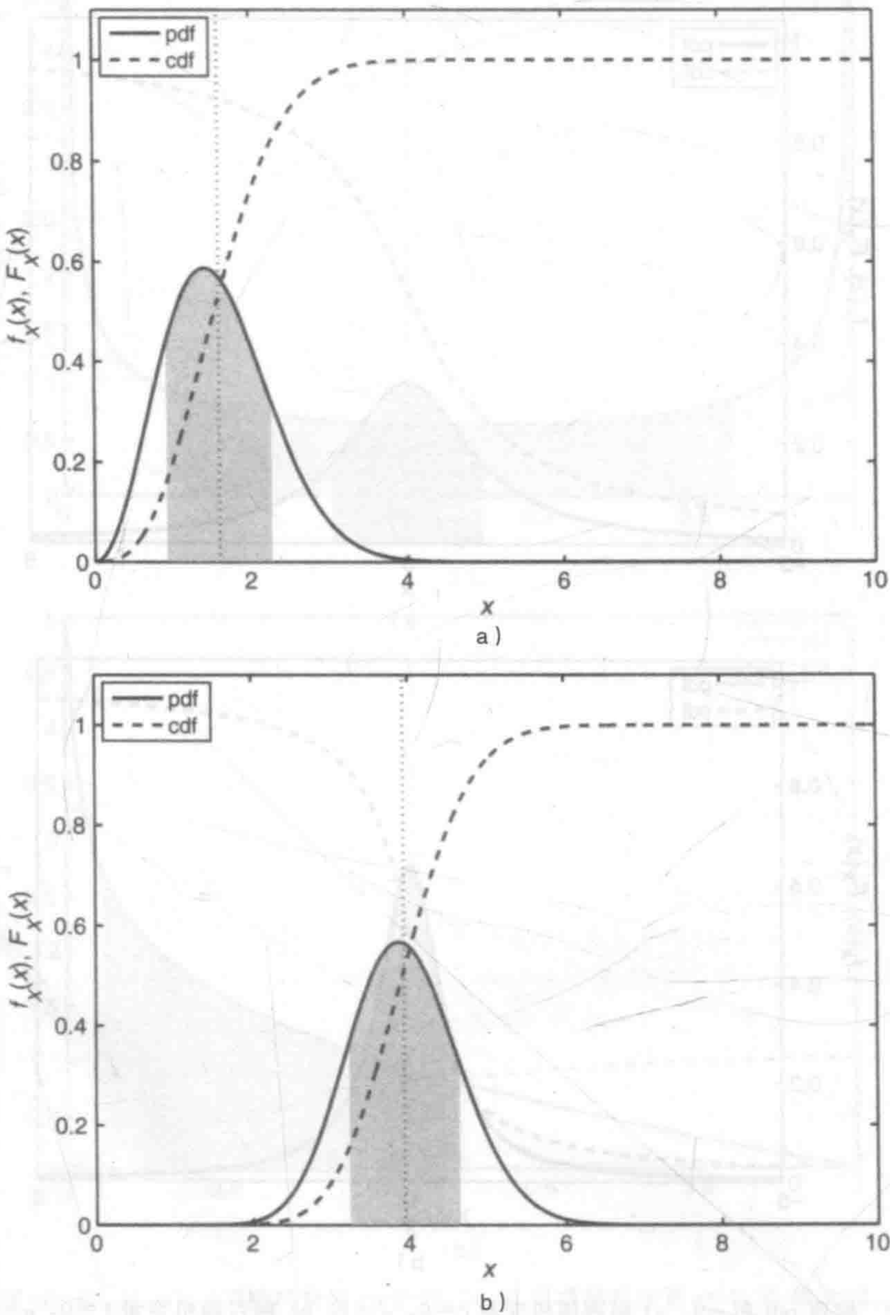


图 A-3 Chi 的 pdf 和 cdf。a) Chi 随机变量 $N=3$；b) Chi 随机变量 $N=16$。虚线表示 μ_X，阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A. 3. 4　卡方

参数 $N \in \mathcal{N}$，定义域 $x \in \mathcal{R}^+$

$$f_X(x) = [x^{N/2-1}/2^{N/2}\Gamma(N/2)]\exp(-x/2)u(x)$$

$$F_X(x) = [\gamma(N/2, x/2)/\Gamma(N/2)]u(x)$$

$$\mu_X = N, \quad \sigma_X^2 = 2N$$

m_e 为隐式，　$m_o = \max(0, N-2)$

$$\mu_n = 2^n\Gamma(n+N/2)/\Gamma(N/2)$$

$$\gamma_1 = \sqrt{8/N}, \quad \gamma_2 = 12/N$$

$$\Phi_X(\omega) = 1/(1-2j\omega)^{N/2}$$

$$h(X) = (1/2)[N + (N-2)\psi(N/2)] + \ln(2\Gamma(N/2))$$

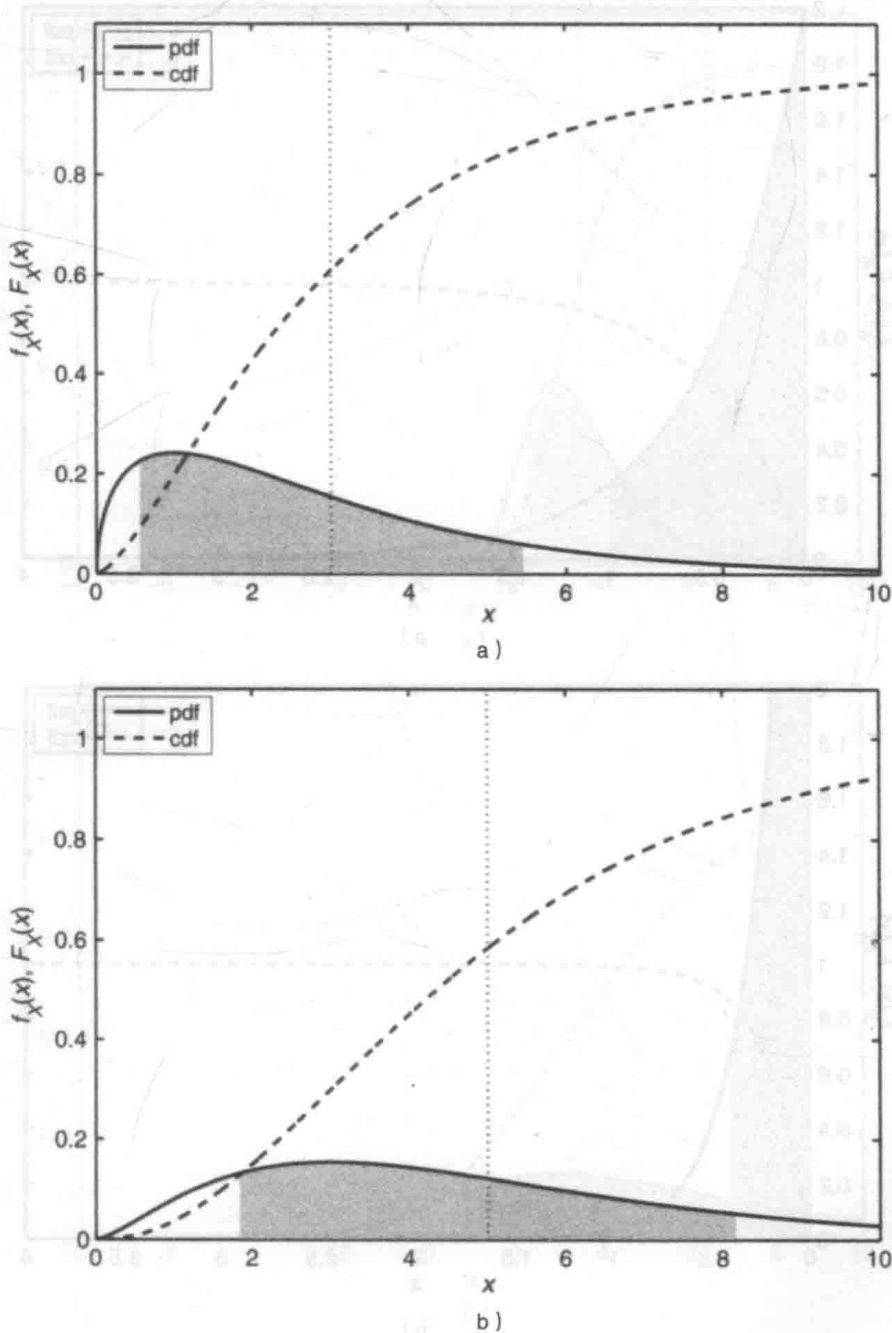


图 A-4　卡方的 pdf 和 cdf。a) 卡方随机变量 $N=3$；b) 卡方随机变量 $N=5$。虚线表示 μ_X，阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A.3.5　指数(平移了 c)

参数 $c \in \mathcal{R}, \lambda > 0$，定义域 $x \in [c, \infty)$

$$f_X(x) = \lambda \exp(-\lambda(x-c))u(x-c)$$

$$F_X(x) = [1-\exp(-\lambda(x-c))]u(x-c)$$

$$\mu_X = c + 1/\lambda, \quad \sigma_X^2 = 1/\lambda^2$$

$$m_e = c + \ln(2)/\lambda, m_o = c$$

$$\mu_n = \Gamma(n+1, \lambda c)\exp(\lambda c)/\lambda^n \quad (\text{对于 } c = 0, \text{为 } n!/\lambda^n)$$

$$\mu_{c,n} = \Gamma(n+1, -1)/e\lambda^n, \quad \gamma_1 = 2, \gamma_2 = 6$$

$$\Phi_X(\omega) = \lambda \exp(j\omega c)/(\lambda - j\omega)$$

$$h(X) = \log(e/\lambda)$$

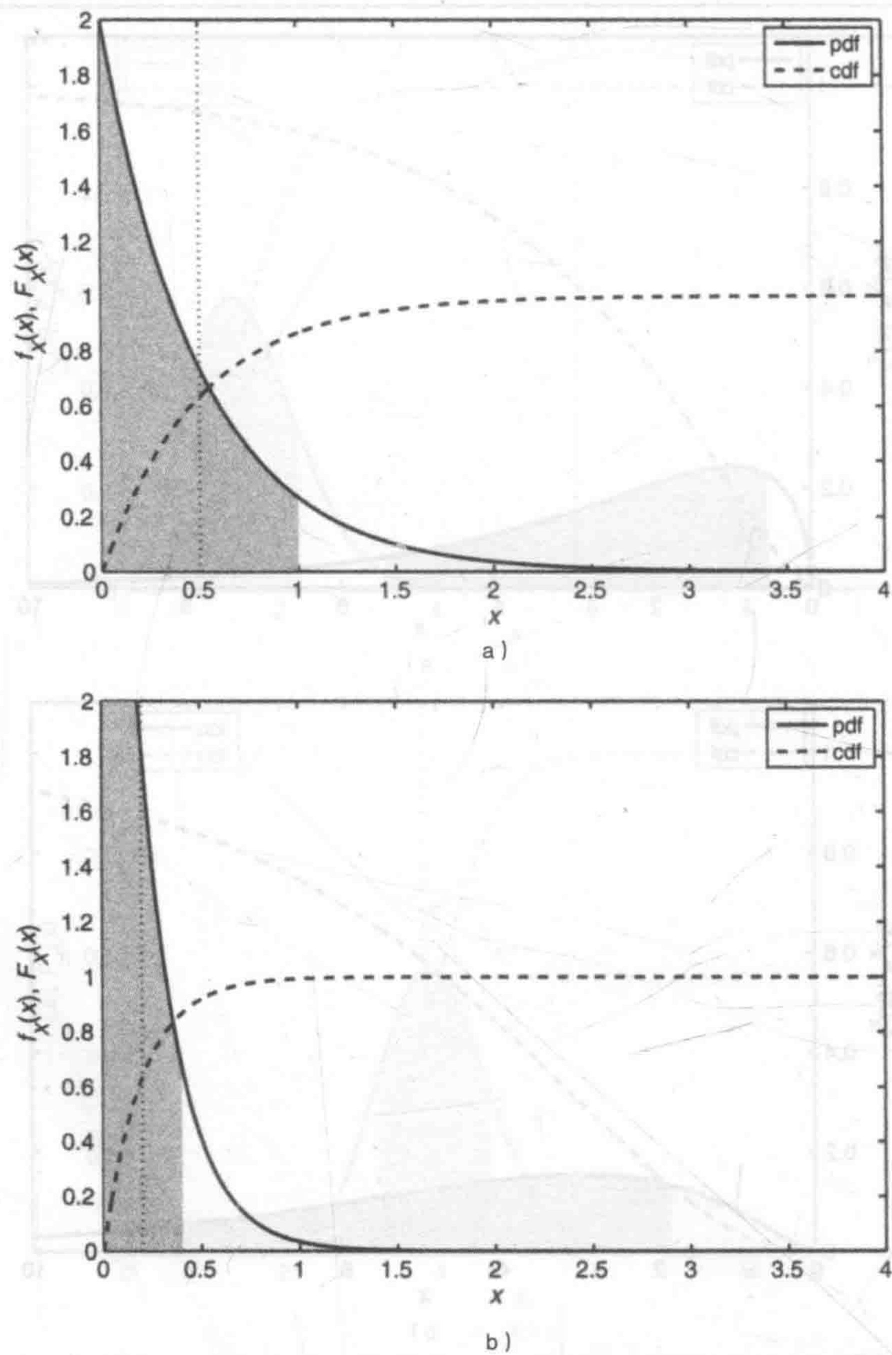


图 A-5　指数 pdf 和 cdf。a) 指数随机变量 $c=0$，$\lambda=2$；b) 指数随机变量 $c=0$，$\lambda=5$。虚线表示 μ_X，阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A.3.6 极值(类型 I：康拜尔)

参数 $c \in \mathcal{R}$，$\alpha > 0$，定义域 $x \in \mathcal{R}$

$$f_X(x) = (1/\alpha)\exp(-(x-c)/\alpha)\exp(-\exp(-(x-c)/\alpha))$$

$$F_X(x) = \exp(-\exp(-(x-c)/\alpha))$$

$$\mu_X = c + \alpha\gamma, \quad \sigma_X^2 = \pi^2\alpha^2/6$$

$$m_e = c - \alpha\ln(\ln(2)), \quad m_o = c$$

μ_n 太复杂

$$\gamma_1 = 12\zeta(3)\sqrt{6}/\pi^3 \approx 1.1395, \quad \gamma_2 = 12/5 = 2.4$$

$$\Phi_X(\omega) = \Gamma(1 - j\omega\alpha)\exp(j\omega c)$$

$$h(X) = \gamma + \ln(\alpha e)$$

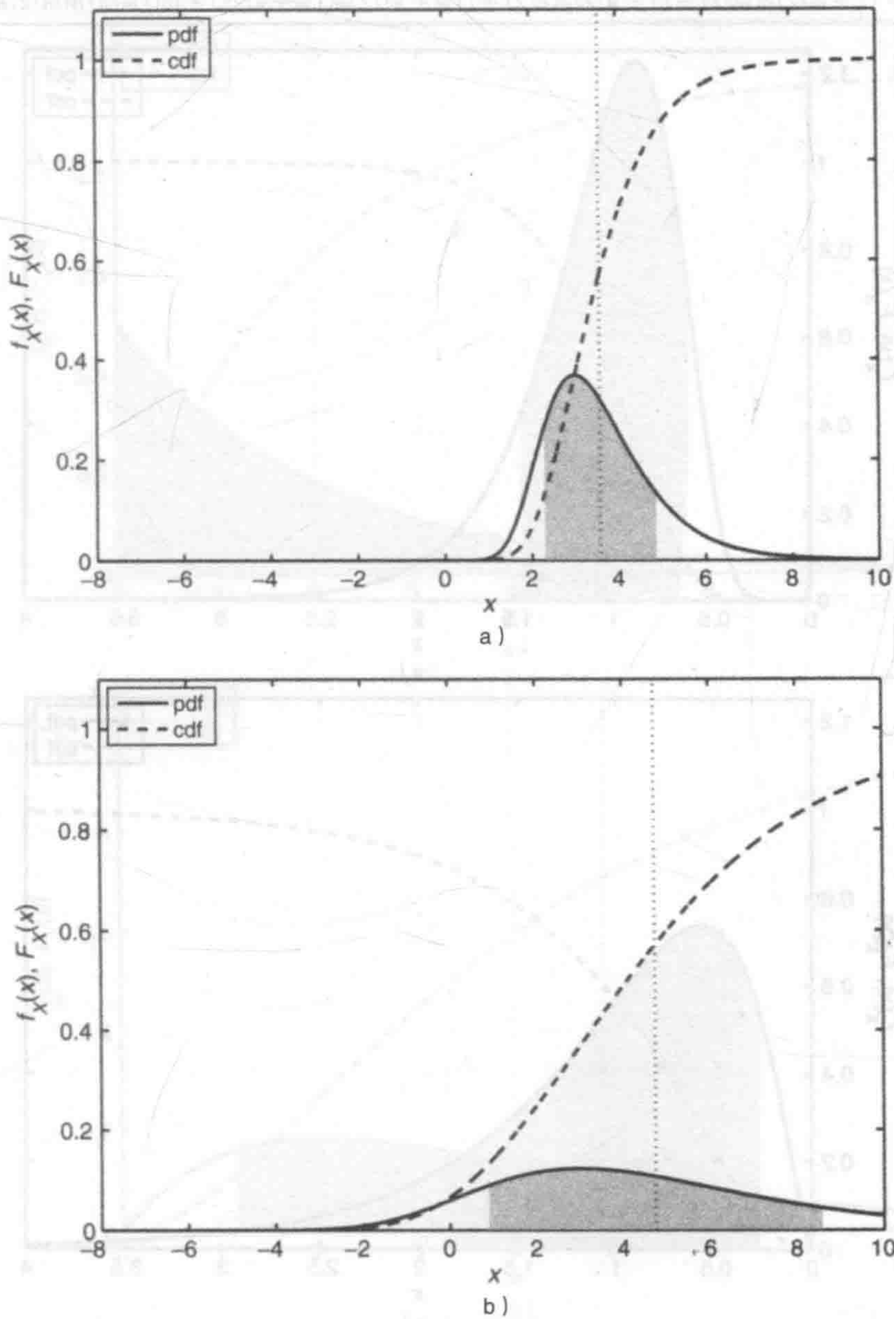


图 A-6 极值(康拜尔)pdf 和 cdf。a) 极值(康拜尔)随机变量 $c=0$，$\alpha=1$；b) 极值(康拜尔)随机变量 $c=3$，$\alpha=3$。虚线表示 μ_X，阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A.3.7　F 分布

参数 $m,n \in \mathcal{N}$，定义域 $x \in \mathcal{R}^+$

$$f_X(x) = \frac{(m/n)^{m/2}x^{(m-2)/2}}{B(m/2,n/2)[1+(m/n)x]^{(m+n)/2}}u(x)$$

$$F_X(x) = [B_{mx/(mx+n)}(m/2,n/2)/B(m/2,n/2)]u(x)$$

$$\mu_X = n/(n-2)(\text{其中 } n>2),\quad \sigma_X^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}(\text{其中 } n>4),m_e \text{ 隐式}$$

$$m_o = n(m-2)/m(n+2)(\text{其中 } m>2),\quad \mu_l = (n/m)^l\frac{\Gamma(m/2+l)\Gamma(n/2-l)}{\Gamma(m/2)\Gamma(n/2)}(\text{其中 } n>2l)$$

$$\gamma_1 - \frac{(2m+n-2)\sqrt{8(n-4)}}{(n-6)\sqrt{m(m+n-2)}}(\text{其中 } n>6),\quad \gamma_2 = \frac{12[(n-2)^2(n-4)+m(m+n-2)(5n-22)]}{m(n-6)(n-8)(m+n-2)}(\text{其中 } n>8)$$

$$\Phi_X(\omega) = [\Gamma((m+n)/2)/\Gamma(n/2)]\Psi(m/2;1-n/2;-\mathrm{j}\omega n/m)$$

$$h(X) = (1-m/2)\psi(m/2)-(1+n/2)\psi(n/2)+[(m+n)/2]\psi((m+n)/2)+\ln((m/n)B(m/2,n/2))$$

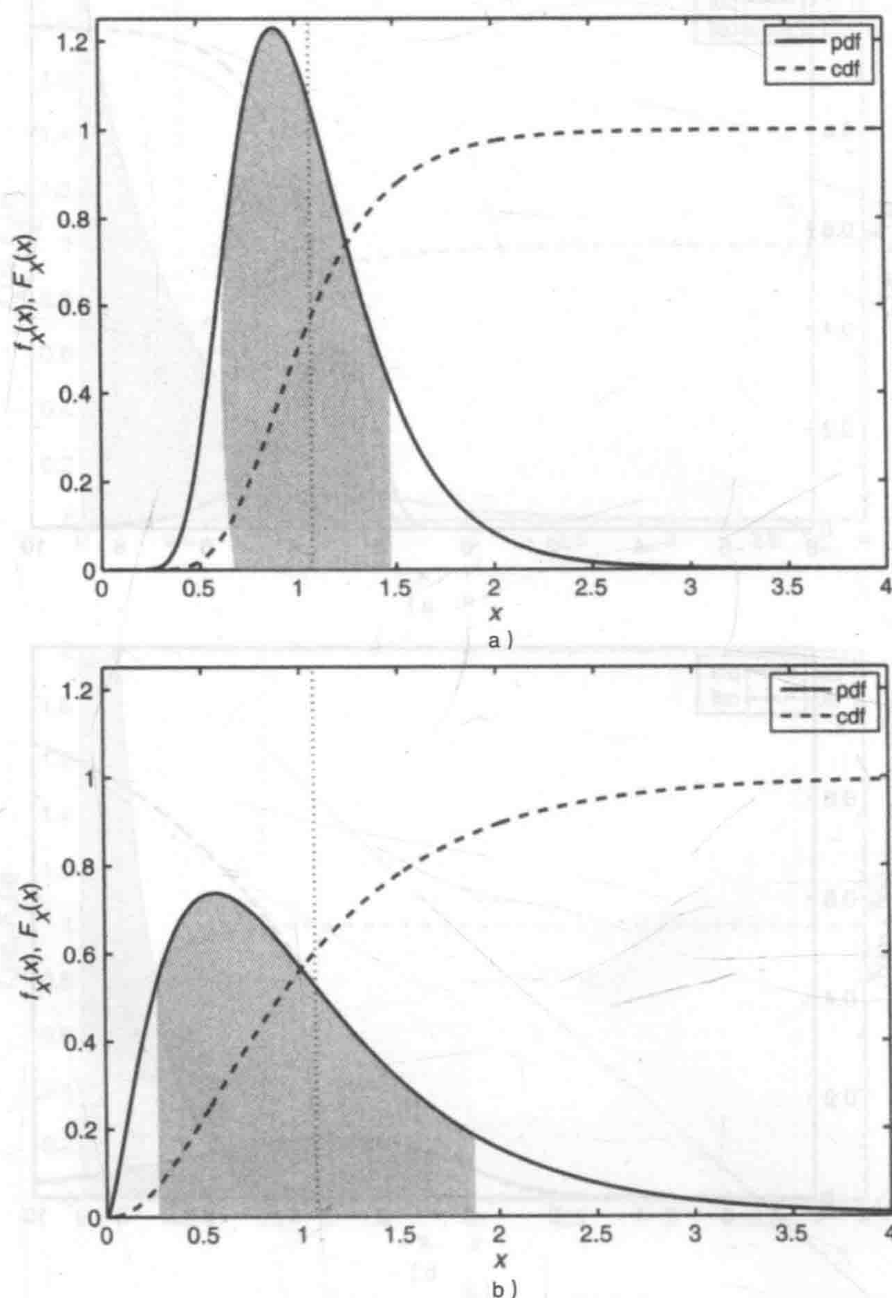


图 A-7　F 分布 pdf 和 cdf。a) F 分布随机变量 $m=40$，$n=30$；b) F 分布随机变量 $m=5$，$n=30$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A.3.8 伽马(埃尔朗，对 $r \in \mathcal{N}$ 有 $\Gamma(r)=(r-1)!$)

$$\text{参数 } \lambda, r > 0, \text{定义域 } x \in \mathcal{R}^+$$
$$f_X(x) = (x^{r-1}\lambda^r/\Gamma(r)]\exp(-\lambda x)u(x)$$
$$F_X(x) = [\gamma(r,\lambda x)/\Gamma(r)]u(x)$$
$$\mu_X = r/\lambda, \quad \sigma_X^2 = r/\lambda^2$$
$$m_e \text{ 隐式}, m_o = (r-1)/\lambda \quad (\text{其中 } r \geqslant 1)$$
$$\mu_n = \Gamma(r+n)/\lambda^n\Gamma(r)$$
$$\gamma_1 = \sqrt{4/r}, \quad \gamma_2 = 6/r$$
$$\Phi_X(\omega) = [\lambda/(\lambda - j\omega)]^r$$
$$h(X) = r + (1-r)\psi(r) + \ln(\Gamma(r)/\lambda)$$

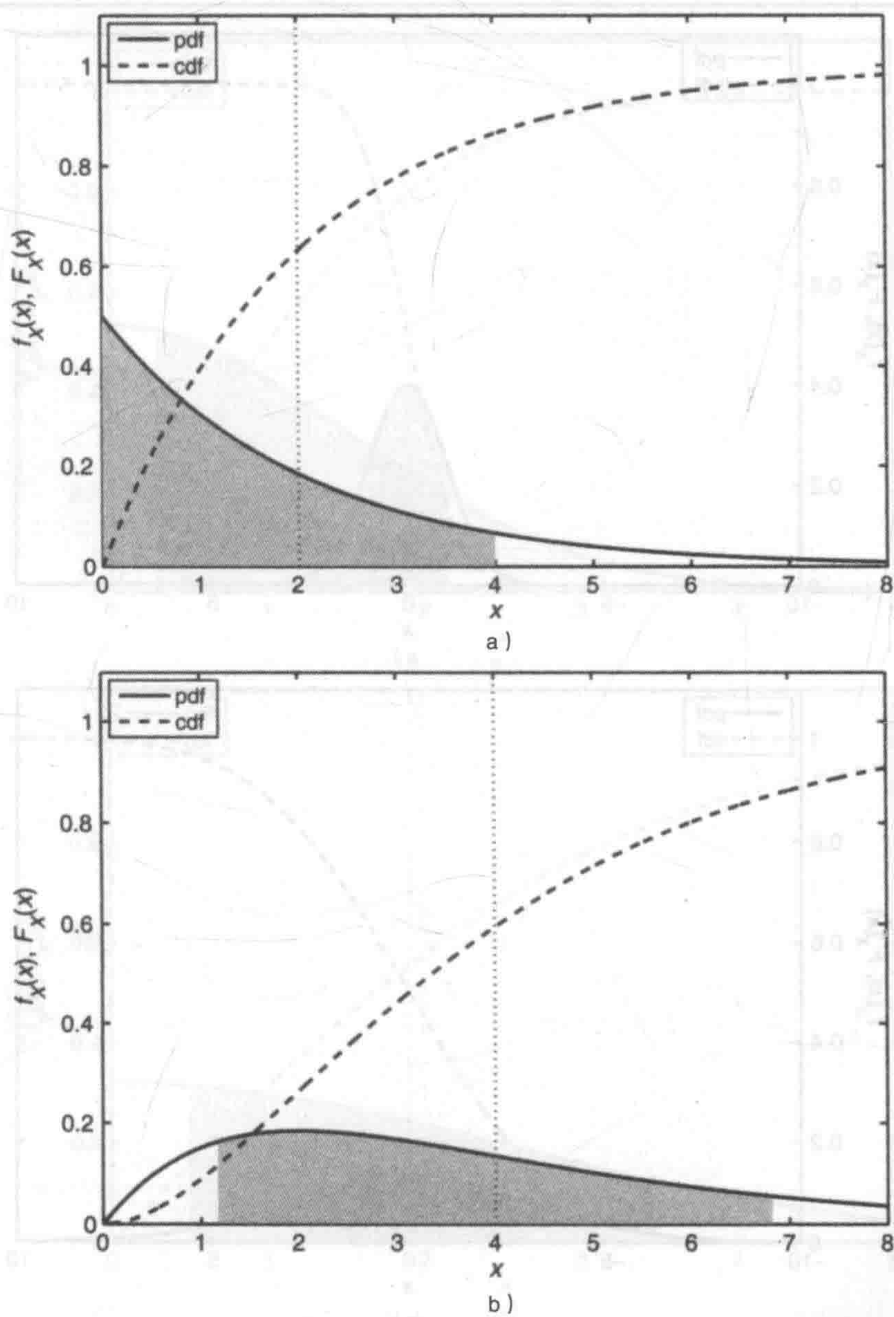


图 A-8 伽马 pdf 和 cdf。a) 伽马随机变量 $\lambda=0.5$，$r=1$；b) 伽马随机变量 $\lambda=0.5$，$r=2$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.3.9　高斯(正态)

参数 $\mu\in\mathcal{R}$，$\sigma>0$，定义域 $x\in\mathcal{R}$

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/2\sigma^2)$$

$$F_X(x)=1-Q\left(\frac{x-\mu}{\sigma}\right)=(1/2)\left[1+\mathrm{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right]$$

$$\mu_X=\mu,\quad \sigma_X^2=\sigma^2,\quad m_e=\mu,\quad m_o=\mu,\quad \mu_2=\sigma^2+\mu^2,\quad \mu_3=\mu(\mu^2+3\sigma^2)$$

$$\gamma_1=0,\quad \gamma_2=0,\quad \mu_{c,n}=\begin{cases}0, & n\text{ 为奇数}\\ \dfrac{\sigma^n n!}{2^{n/2}(n/2)!}, & n\text{ 为偶数}\end{cases}$$

$$\Phi_X(\omega)=\exp(\mathrm{j}\omega\mu-\sigma^2\omega^2/2),\quad h(X)=(1/2)\log(2\pi\sigma^2\mathrm{e})$$

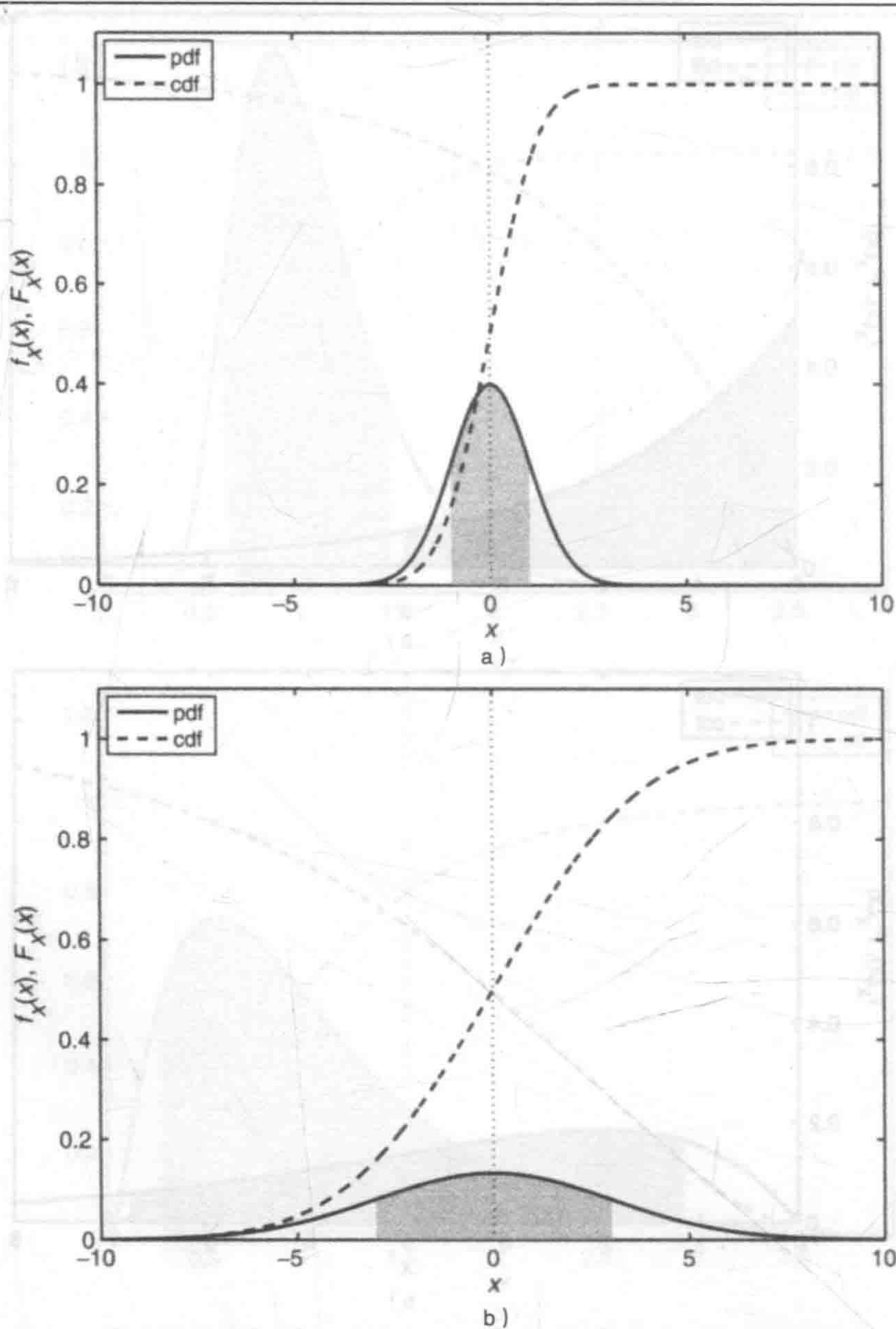


图 A-9　高斯 pdf 和 cdf。a）高斯随机变量 $\mu=0$，$\sigma=1$；b）高斯随机变量 $\mu=0$，$\sigma=3$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A.3.10 半正态(折叠正态)

参数 $\alpha>0$,定义域 $x\in\mathcal{R}^+$

$$f_X(x)=\sqrt{2/\pi\alpha^2}\exp(-x^2/2\alpha^2)u(x)$$

$$F_X(x)=\mathrm{erf}(x/\alpha\sqrt{2})u(x)$$

$$\mu_X=\alpha\sqrt{2/\pi},\quad \sigma_X^2=\alpha^2(\pi-2)/\pi,\quad m_e\text{ 为隐式},\quad m_o=0$$

$$\mu_n=(2^{n/2}\alpha^n/\sqrt{\pi})\Gamma((n+1)/2)$$

$$\gamma_1=\sqrt{2}(4-\pi)/(\pi-2)^{3/2}\approx 0.9953$$

$$\gamma_2=8(\pi-3)/(\pi-2)^2\approx 0.8692$$

$$\Phi_X(\omega)=[1+\mathrm{j}\,\mathrm{sgn}(\omega)\mathrm{erfi}(\alpha|\omega|/\sqrt{2})]\exp(-\omega^2\alpha^2/2)$$

$$h(X)=(1/2)\log(\pi\alpha^2\mathrm{e}/2)$$

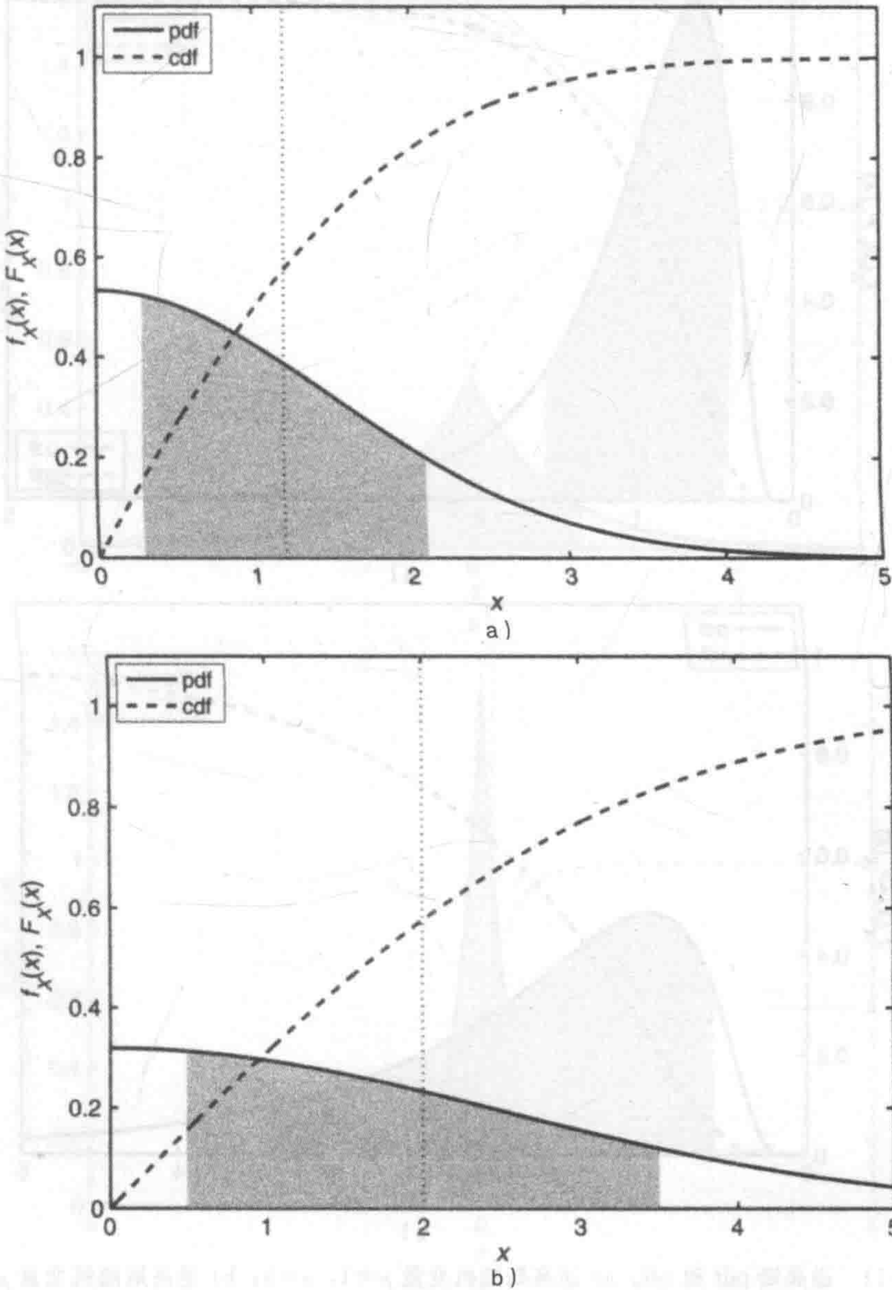


图 A-10 半正态 pdf 和 cdf。a) 半正态随机变量 $\alpha=1.5$;b) 半正态随机变量 $\alpha=2.5$。虚线表示 μ_X,阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A. 3. 11 逆高斯(Wald)

参数 $\mu, \alpha > 0$，定义域 $x \in \mathcal{R}^+$

$$f_X(x) = \sqrt{\frac{\alpha}{2\pi x^3}} \exp(-\alpha(x-\mu)^2/2\mu^2 x)u(x)$$

$$F_X(x) = (1 - Q((x/\mu - 1)\ \sqrt{\alpha/x}) + \exp(2\alpha/\mu)(Q((x/\mu + 1)\ \sqrt{\alpha/x})])u(x)$$

$$\mu_X = \mu, \quad \sigma_X^2 = \mu^3/\alpha, \quad m_e \text{ 为隐式}, \quad m_o = \mu\ \sqrt{1 + (3\mu/2\alpha)^2} - 3\mu^2/2\alpha$$

$$\mu_2 = \mu^2(1 + \mu/\alpha), \quad \mu_3 = \mu^3(1 + 3\mu/\alpha + 3^2/\alpha^2), \quad \mu_{c,3} = 3\mu^5/\alpha^2$$

$$\gamma_1 = 3\ \sqrt{\mu/\alpha}, \quad \gamma_2 = 15\mu/\alpha$$

$$\Phi_X(\omega) = \exp((\alpha/\mu)[1 - \sqrt{1 - 2j\omega\mu^2/\alpha}]), \quad h(X) \text{ 太复杂}$$

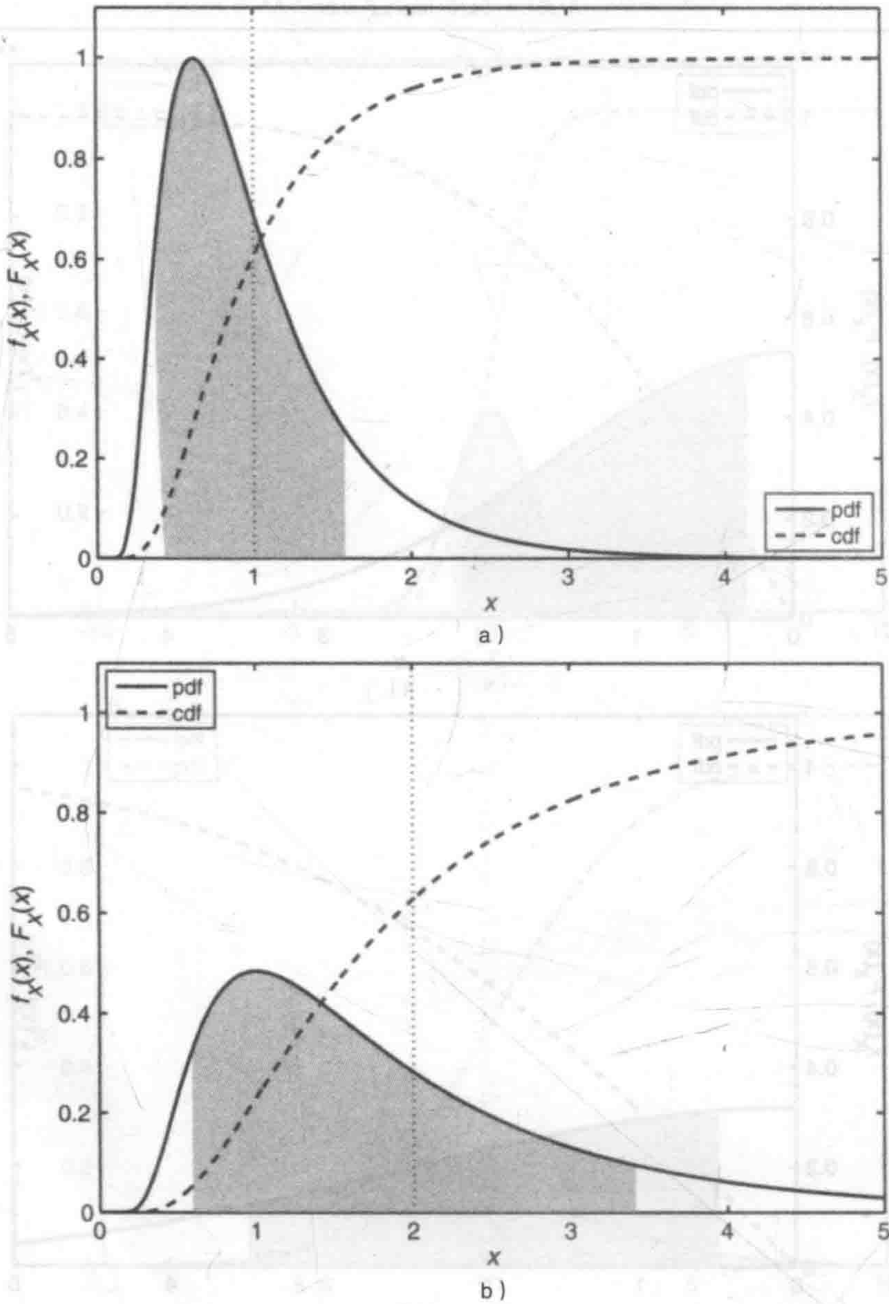


图 A-11 逆高斯 pdf 和 cdf。a) 逆高斯随机变量 $\mu=1$，$\alpha=3$；b) 逆高斯随机变量 $\mu=2$，$\alpha=4$。虚线表示 μ_X，阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A.3.12 拉普拉斯(双边指数)

参数 $\mu \in \mathcal{R}, \alpha > 0$，定义域 $x \in \mathcal{R}$

$$f_X(x) = (1/2\alpha)\exp(-|x-\mu|/\alpha)$$

$$F_X(x) = (1/2)(1 + \mathrm{sgn}(x-\mu)[1-\exp(-|x-\mu|/\alpha)])$$

$$\mu_X = \mu, \quad \sigma_X^2 = 2\alpha^2, \quad m_e = \mu, m_o = \mu$$

$$\mu_{c,n} = \begin{cases} 0, & n\text{ 为奇数} \\ \alpha^n n!, & n\text{ 为偶数,} \end{cases}$$

$$\gamma_1 = 0, \quad \gamma_2 = 3$$

$$\Phi_X(\omega) = \exp(j\omega\mu)/(1+\alpha^2\omega^2)$$

$$h(X) = \log(2\alpha e)$$

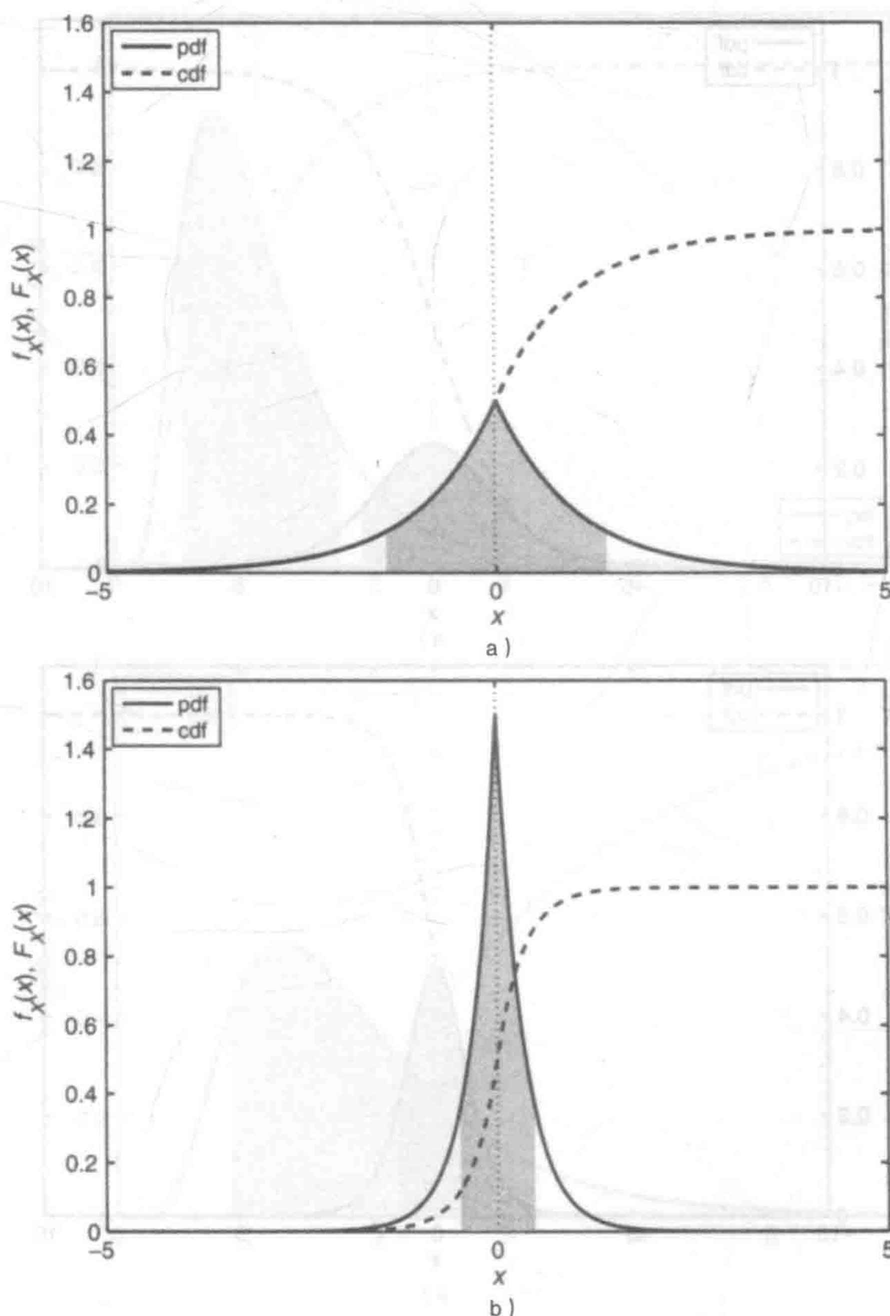


图 A-12 拉普拉斯 pdf 和 cdf。a) 拉普拉斯随机变量 $\mu=0$，$\alpha=1$；b) 拉普拉斯随机变量 $\mu=0$，$\alpha=1/3$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.3.13 Logistic(参数为$\{\mu=0,\ \alpha=1\}$的 Sigmoid)

参数 $\mu \in \mathcal{R}, \alpha > 0$,定义域 $x \in \mathcal{R}^+$

$$f_X(x) = (1/\alpha)\exp(-(x-\mu)/\alpha)[1+\exp(-(x-\mu)/\alpha)]^{-2} = (1/4\alpha)\operatorname{sech}^2((x-\mu)/2\alpha)$$

$$F_X(x) = [1+\exp(-(x-\mu)/\alpha)]^{-1} = (1/2)[1+\tanh((x-\mu)/2\alpha)]$$

$$\mu_X = \mu, \quad \sigma_X^2 = \pi^2\alpha^2/3$$

$$m_e = \mu, \quad m_o = \mu,$$

$$\mu_{c,n} = (\alpha\pi)^n(2^n-2)\,|B_n|$$

$$\gamma_1 = 0, \quad \gamma_2 = 6/5$$

$$\Phi_X(\omega) = \omega\pi\alpha\exp(\mathrm{j}\omega\mu)/\sinh(\omega\pi\alpha)$$

$$h(X) = 2+\ln(\alpha)$$

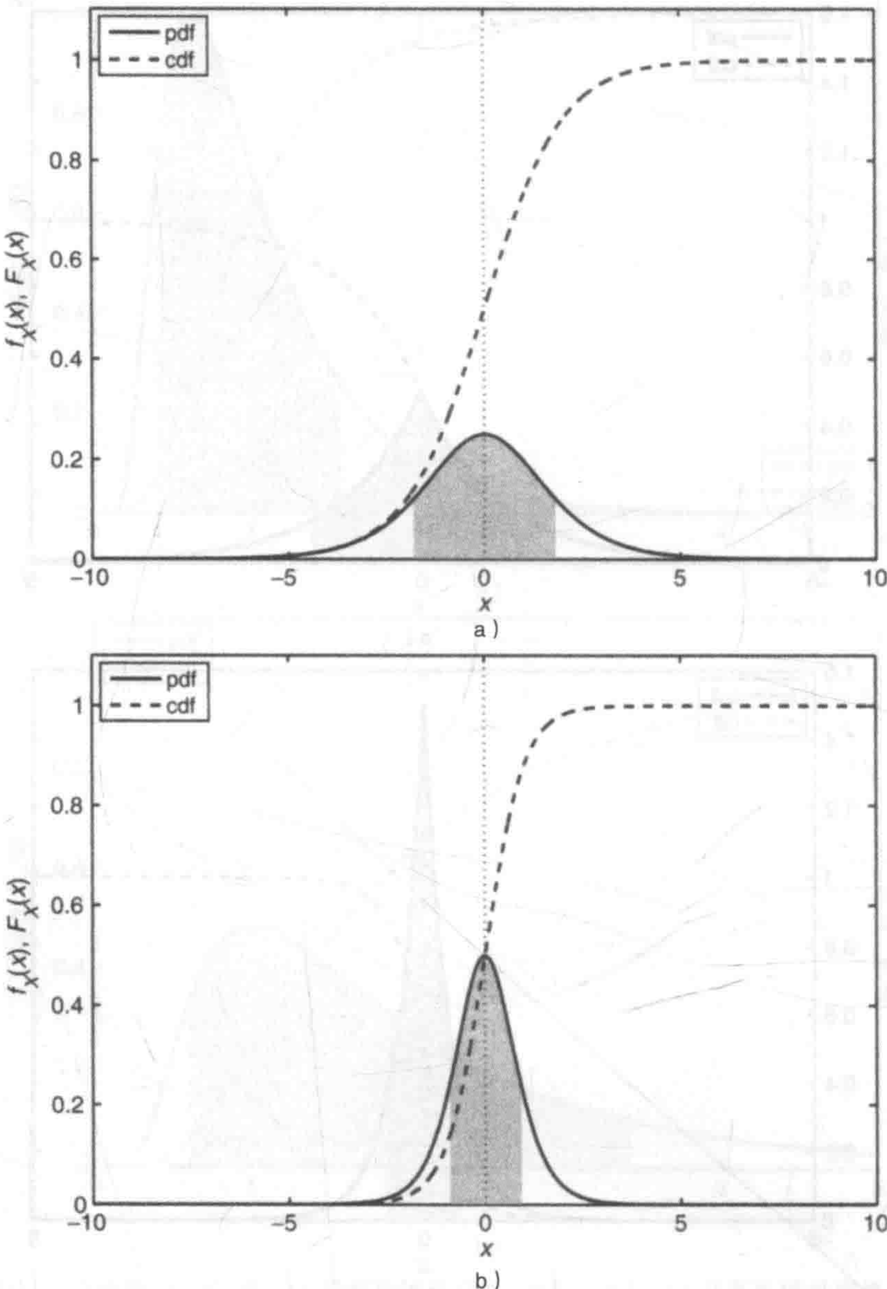


图 A-13 Logistic pdf 和 cdf。a) Logistic 随机变量 $\mu=0$,$\alpha=1$;b) Logistic 随机变量$\mu=0$,$\alpha=0.5$。虚线表示 μ_X,阴影部分表示$[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.3.14 对数正态

参数 $\mu \in \mathcal{R}, \sigma > 0$，定义域 $x \in \mathcal{R}^+$

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\exp(-(\ln(x)-\mu)^2/2\sigma^2)u(x)$$

$$F_X(x) = \left[1 - Q\left(\frac{\ln(x)-\mu}{\sigma}\right)\right]u(x) = (1/2)\left[1 + \mathrm{erf}\left(\frac{\ln(x)-\mu}{\sqrt{2\sigma^2}}\right)\right]u(x)$$

$$\mu_X = \exp(\mu + \sigma^2/2), \quad \sigma_X^2 = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$$

$$m_e = \exp(\mu), \quad m_o = \exp(\mu - \sigma^2), \quad \mu_n = \exp(n\mu + n^2\sigma^2/2)$$

$$\gamma_1 = [\exp(\sigma^2) + 2]\sqrt{\exp(\sigma^2) - 1}, \quad \gamma_2 = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 6$$

$\Phi_X(\omega)$ 太复杂， $h(X) = \mu + (1/2)\ln(2\pi\sigma^2 e)$

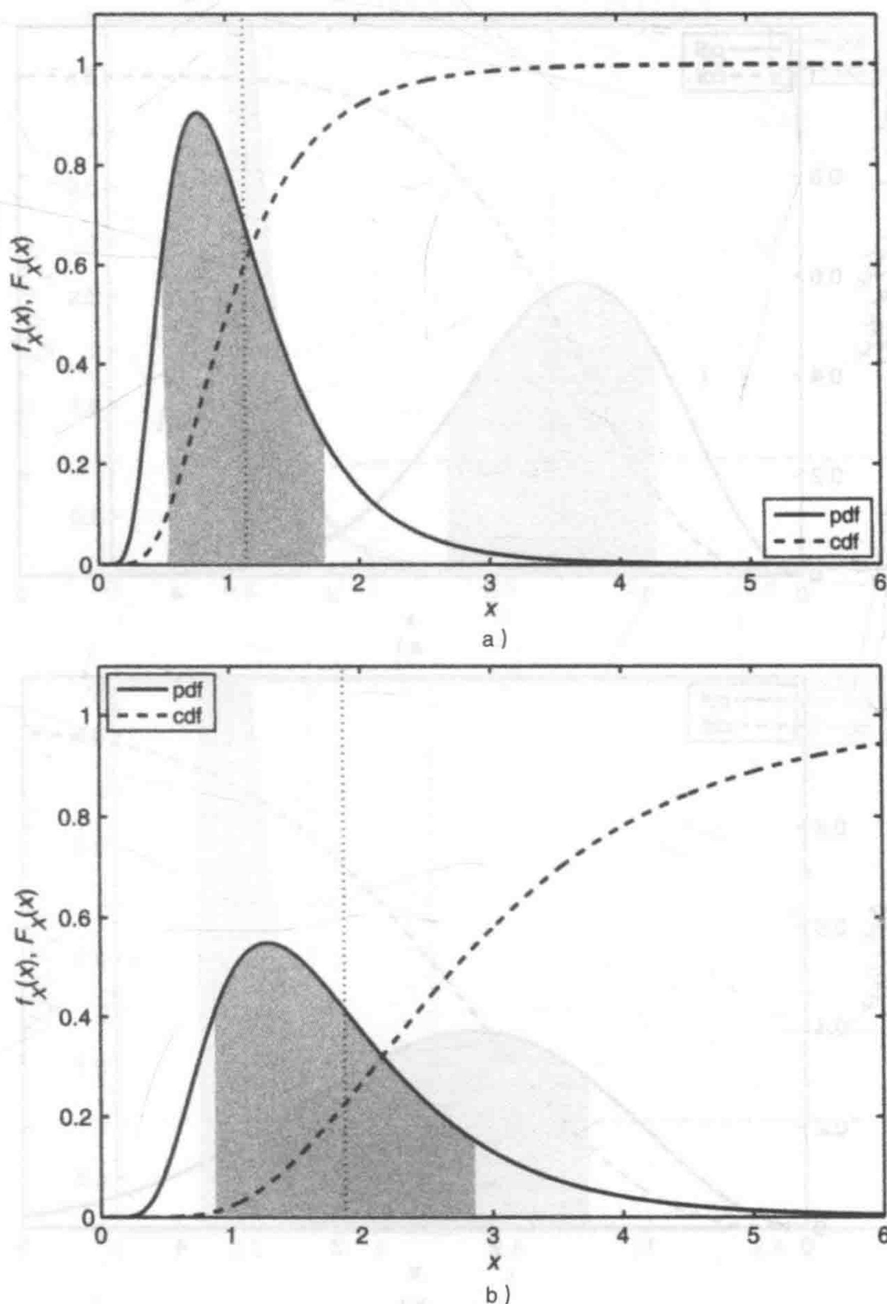


图 A-14 对数正态 pdf 和 cdf。a) 对数正态随机变量 $\mu=0$，$\sigma=0.5$；b) 对数正态随机变量 $\mu=0.5$，$\sigma=0.5$。虚线表示 μ_X，阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A.3.15 麦克斯韦-玻耳兹曼

参数 $\alpha > 0$，定义域 $x \in \mathcal{R}^+$

$$f_X(x) = (1/\alpha^3)\sqrt{2/\pi}x^2\exp(-x^2/2\alpha^2)u(x)$$

$$F_X(x) = [\mathrm{erf}(x/\alpha\sqrt{2}) - \sqrt{2/\pi}(x/\alpha)\exp(-x^2/2\alpha^2)]u(x)$$

$$\mu_X = 2\alpha\sqrt{2/\pi}, \quad \sigma_X^2 = \alpha^2(3\pi - 8)/\pi$$

m_e 为隐式， $m_o = \alpha\sqrt{2}$， $\mu_n = [2^{(n+2)/2}\alpha^n/\sqrt{\pi}]\Gamma((n+3)/2)$

$$\gamma_1 = 2\sqrt{2}(5\pi - 16)/(3\pi - 8)^{3/2} \approx -0.4857$$

$$\gamma_2 = 4(3\pi^2 - 40\pi + 96)/(3\pi - 8)^2 \approx -0.1082$$

$$\Phi_X(\omega) = \mathrm{j}\omega\alpha\sqrt{2/\pi} + (1 - \omega^2\alpha^2)[1 + \mathrm{jsgn}(\omega)\mathrm{erfi}(\alpha|\omega|/\sqrt{2})]\exp(-\omega^2\alpha^2/2)$$

$$h(X) = \gamma + (1/2)[\ln(2\pi\alpha^2) - 1]$$

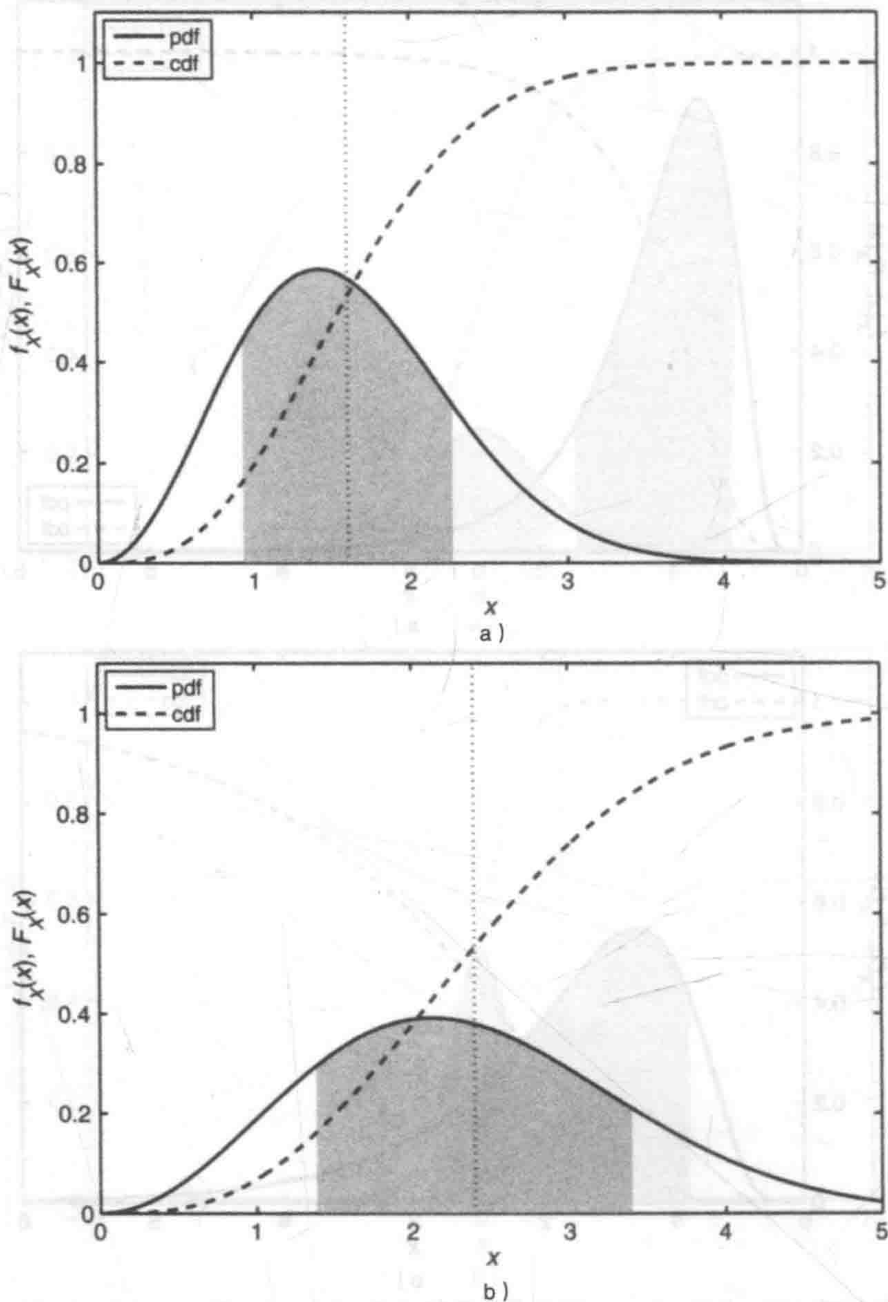


图 A-15 麦克斯韦-玻耳兹曼 pdf 和 cdf。a) 麦克斯韦-玻耳兹曼随机变量 $\alpha=1$；b) 麦克斯韦-玻耳兹曼随机变量 $\alpha=1.5$。虚线表示 μ_X，阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A. 3. 16　帕雷托

参数 $r, \alpha > 0$，定义域 $x \in [\alpha, \infty)$

$$f_X(x) = [r\alpha^r / x^{r+1}]u(x-\alpha)$$

$$F_X(x) = [1-(\alpha/x)^r]u(x-\alpha)$$

$$\mu_X = r\alpha/(r-1)（其中\ r>1）, \sigma_X^2 = r\alpha^2/[(r-1)^2(r-2)]（其中\ r>2）$$

$$m_e = \alpha 2^{1/r}, \quad m_o = \alpha, \quad \mu_n = r\alpha^n/(r-n)（其中\ r>n）$$

$$\gamma_1 = [2(r+1)/(r-3)]\sqrt{(r-2)/r}（其中\ r>3）, \quad \gamma_2 = \frac{6(r^3+r^2-6r-2)}{r(r-3)(r-4)}（其中\ r>4）$$

$$\Phi_X(\omega) = r(-\mathrm{j}\omega\alpha)^r\Gamma(-r, -\mathrm{j}\omega\alpha)$$

$$h(X) = (r+1)/r + \ln(\alpha/r)$$

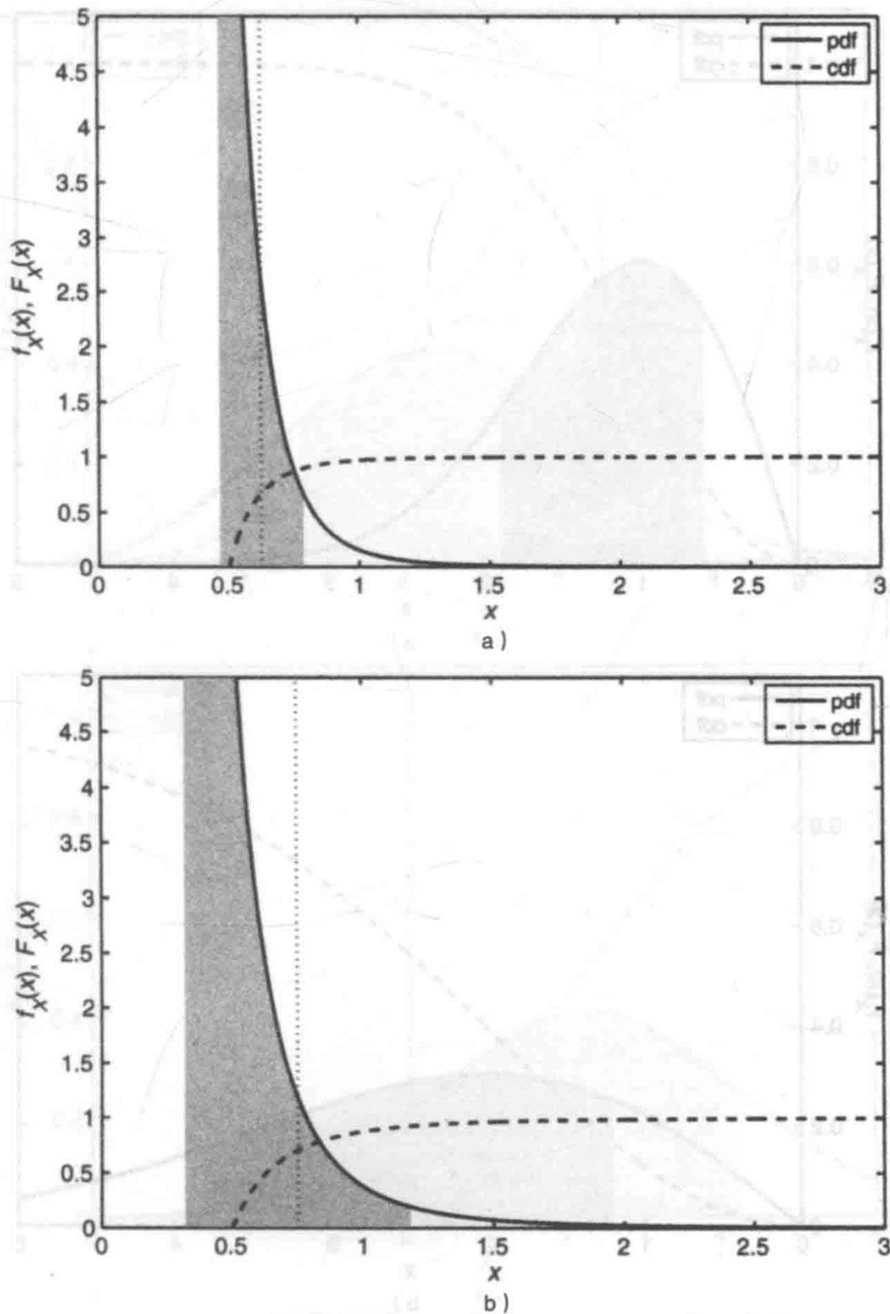


图 A-16　帕雷托 pdf 和 cdf。a）帕雷托随机变量 $\alpha=0.5$，$r=5$；b）帕雷托随机变量 $\alpha=0.5$，$r=3$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.3.17 瑞利

参数 $\alpha > 0$，定义域 $x \in \mathcal{R}^+$

$$f_X(x) = (x/\alpha^2)\exp(-x^2/2\alpha^2)u(x)$$

$$F_X(x) = [1-\exp(-x^2/2\alpha^2)]u(x)$$

$$\mu_X = \alpha\sqrt{\pi/2},\quad \sigma_X^2 = (4-\pi)\alpha^2/2,\quad m_e = \alpha\sqrt{\ln(4)},m_o = \alpha$$

$$\mu_n = 2^{n/2}\alpha^n\Gamma(n/2+1),\quad \mu_{c,n} = \sqrt{\pi/2}(\pi-3)\alpha^3$$

$$\gamma_1 = 2(\pi-3)\sqrt{\pi}/(4-\pi)^{3/2} \approx 0.6311,\quad \gamma_2 = (24\pi-6\pi^2-16)/(4-\pi)^2 \approx 0.2451$$

$$\Phi_X(\omega) = 1-\omega\alpha\sqrt{\pi/2}\exp(-\alpha^2\omega^2/2)[\mathrm{erfi}(\omega\alpha/\sqrt{2})-\mathrm{j}]$$

$$h(X) = \gamma/2+1+\ln(\alpha/\sqrt{2})$$

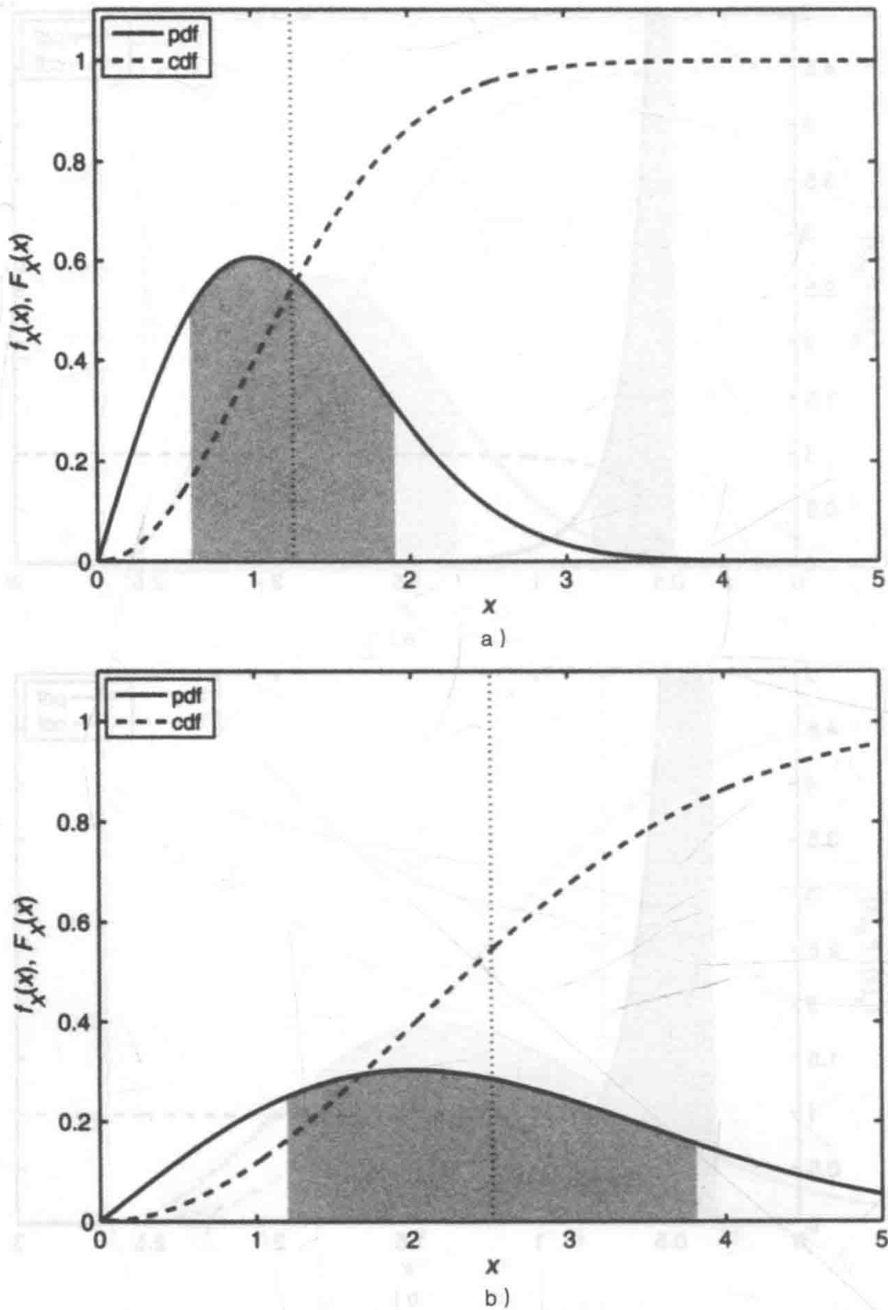


图 A-17 瑞利 pdf 和 cdf。a) 瑞利随机变量 $\alpha=1$；b) 瑞利随机变量 $\alpha=2$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.3.18 莱斯

参数 $c,\alpha>0$，定义域 $x\in\mathcal{R}^+$

$$f_X(x)=(x/\alpha^2)\exp(-(x^2+c^2)/2\alpha^2)I_0(cx/\alpha^2)u(x)$$

$$F_X(x)=[1-Q_1(c/\alpha,x/\alpha)]u(x)$$

$$\mu_X=\alpha\sqrt{\pi/2}L_{1/2}(-c^2/2\alpha^2),\sigma_X^2=c^2+2\alpha^2-(\pi\alpha^2/2)L_{1/2}^2(-c^2/2\alpha^2)$$

m_e 为隐式， m_o 为隐式， γ_1 太复杂， γ_2 太复杂

$$\mu_n=(2\alpha^2)^{n/2}\Gamma(n/2+1)L_{n/2}(-c^2/2\alpha)$$

$$\Phi_X(\omega)=\sum_{n=0}^{\infty}[(\mathrm{j}\omega\alpha\sqrt{2})^n/n!]\Gamma(n/2+1)L_{n/2}(-c^2/2\alpha)$$

$h(X)$ 太复杂

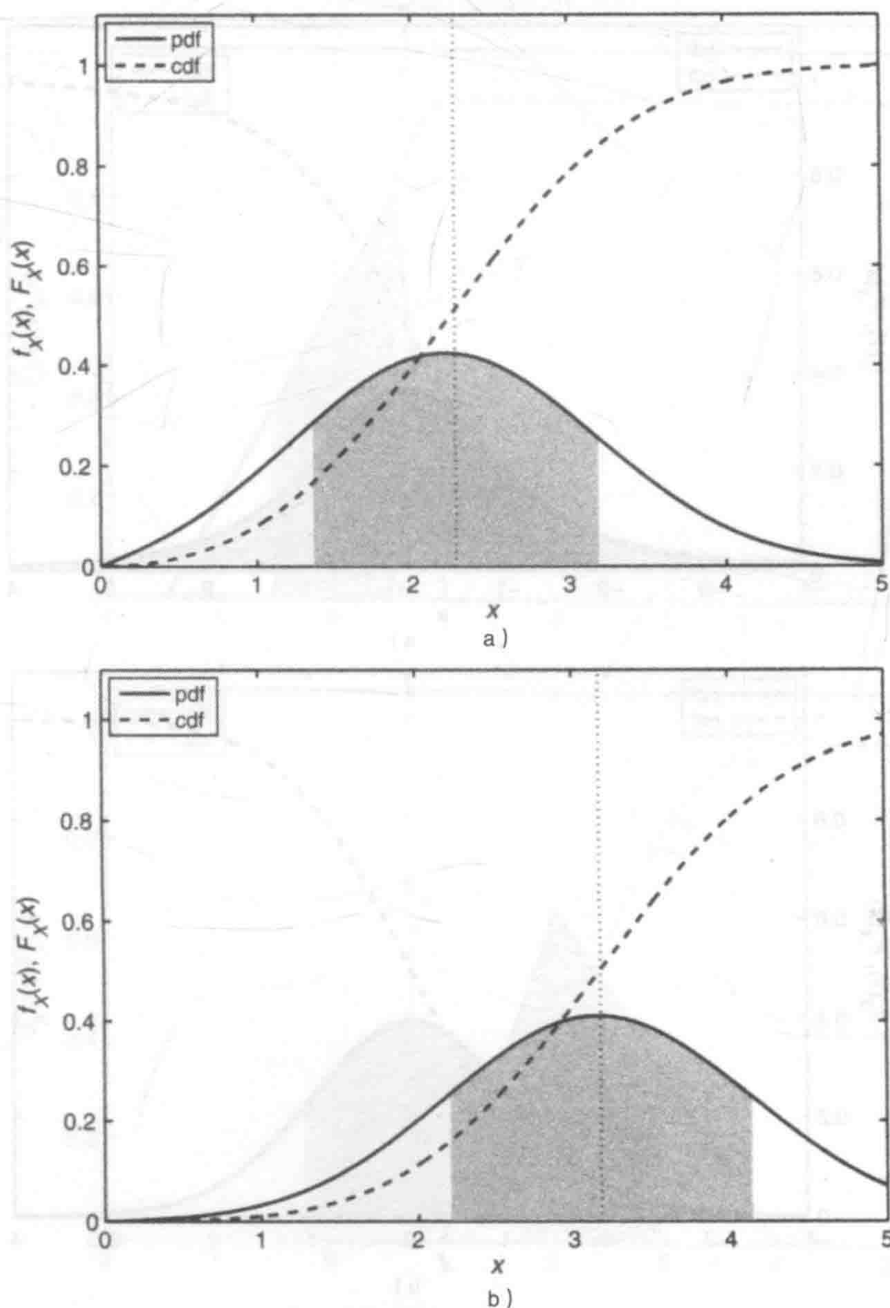


图 A-18 莱斯 pdf 和 cdf。a) 莱斯随机变量 $c=2$ and $\alpha=1$；b) 莱斯随机变量 $c=3$ and $\alpha=1$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A.3.19 学生氏 t 分布

参数 $r>0$，定义域 $x\in\mathcal{R}^{+}$

$$f_X(x)=(1+x^2/r)^{-(r+1)/2}/\sqrt{r}B(1/2,r/2)$$

$$F_X(x)=B_{(x+\sqrt{x^2+r})/2\sqrt{x^2+r}}(r/2,r/2)/B(r/2,r/2)$$

$$\mu_X=0(其中\ r>1),\sigma_X^2=r/(r-2)(其中\ r>2),m_e=0,m_o=0$$

$$\mu_{c,n}=\begin{cases}0, & n\ 奇数\\ r^{n/2}B((n+1)/2,(r-n)/2)/B(1/2,k/2), & n\ 偶数\end{cases}(其中\ r>n)$$

$$\gamma_1=0(其中\ r>3),\quad \gamma_2=6/(r-4)(其中\ r>4)$$

$$\Phi_X(\omega)=K_{r/2}(\sqrt{r}|\omega|)\sqrt{(\sqrt{r}|\omega|)^r}/\Gamma(r/2)2^{r/2-1}$$

$$h(X)=[\psi((r+1)/2)-\psi(r/2)](r+1)/2+\ln(\sqrt{r}B(1/2,r/2))$$

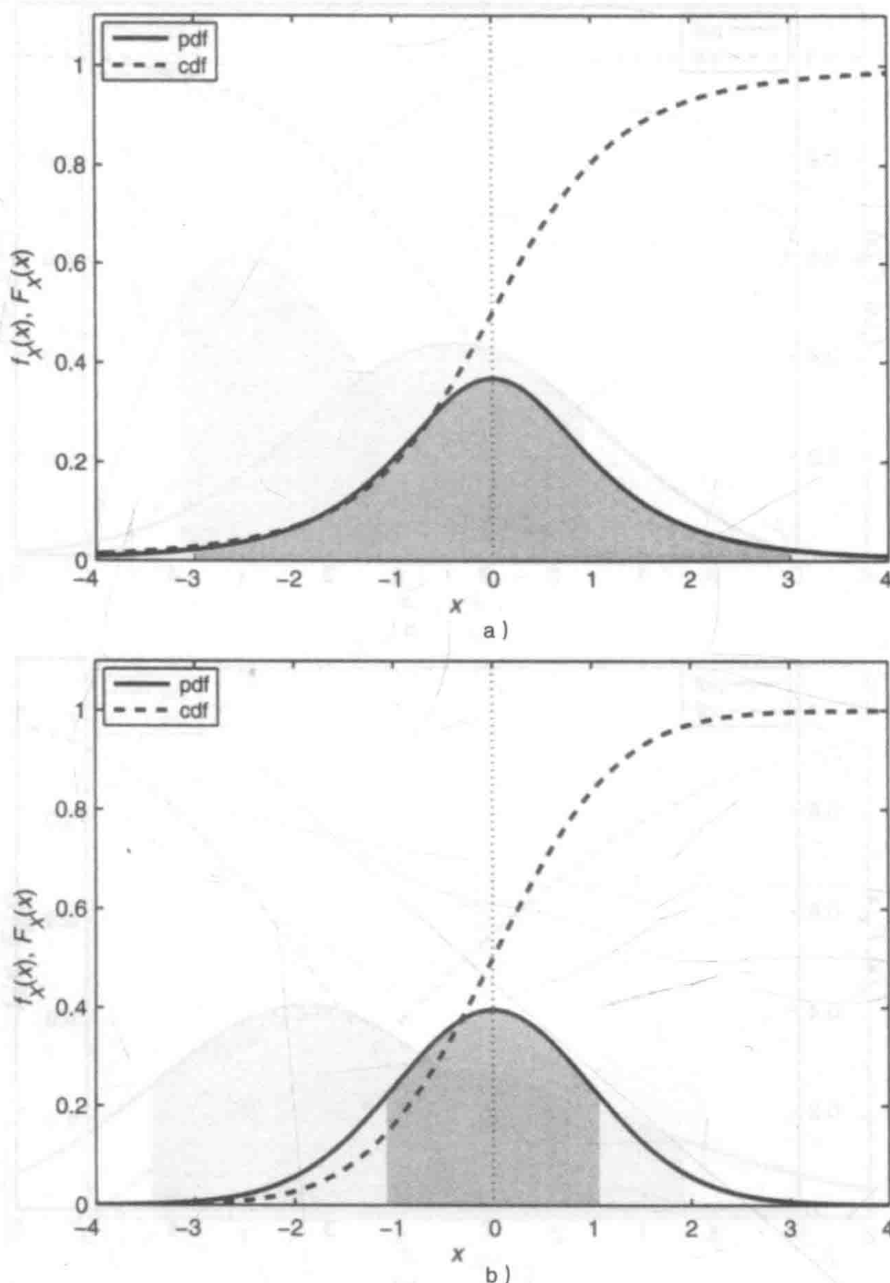


图 A-19 学生氏 t 分布 pdf 和 cdf。a) 学生氏 t 分布随机变量 $r=3$；b) 学生氏 t 分布随机变量 $r=30$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A. 3. 20 三角

参数 $a,b,c \in \mathcal{R}$（其中 $a<c<b$），定义域 $x \in [a,b]$

$$f_X(x)=[2/(b-a)]\left[\frac{x-a}{c-a}I_{(a,c]}(x)+\frac{b-x}{b-c}I_{(c,b]}(x)\right]$$

$$F_X(x)=[1/(b-a)]\left[\frac{(x-a)^2}{c-a}I_{(a,c]}(x)+\left(b-a-\frac{(b-x)^2}{b-c}\right)I_{(c,b]}(x)\right]+I_{(b,\infty)}(x)$$

$$\mu_X=(a+b+c)/3,\quad \sigma_X^2=(a^2+b^2+c^2-ab-ac-bc)/18$$

$$m_o=c,\quad m_e=\begin{cases}b-\sqrt{(b-a)(-c)/2}, & c\leqslant(a+b)/2\\ a+\sqrt{(b-a)(c-a)/2}, & c>(a+b)/2\end{cases}$$

$$\mu_2=(a^2+b^2+c^2+ab+ac+bc)/6,\quad \gamma_1=\mu_{c,3}/\sigma_X^3,\quad \gamma_2=-3/5=-0.6$$

$$\mu_{c,3}=-(a+b-2c)(a+c-2b)(b+c-2a)/270, h(X)=1/2+\log((b-a)/2)$$

$$\Phi_X(\omega)=2\frac{(b-a)\exp(\mathrm{j}\omega c)-(b-c)\exp(\mathrm{j}\omega a)-(c-a)\exp(\mathrm{j}\omega b)}{\omega^2(b-a)(b-c)(c-a)}$$

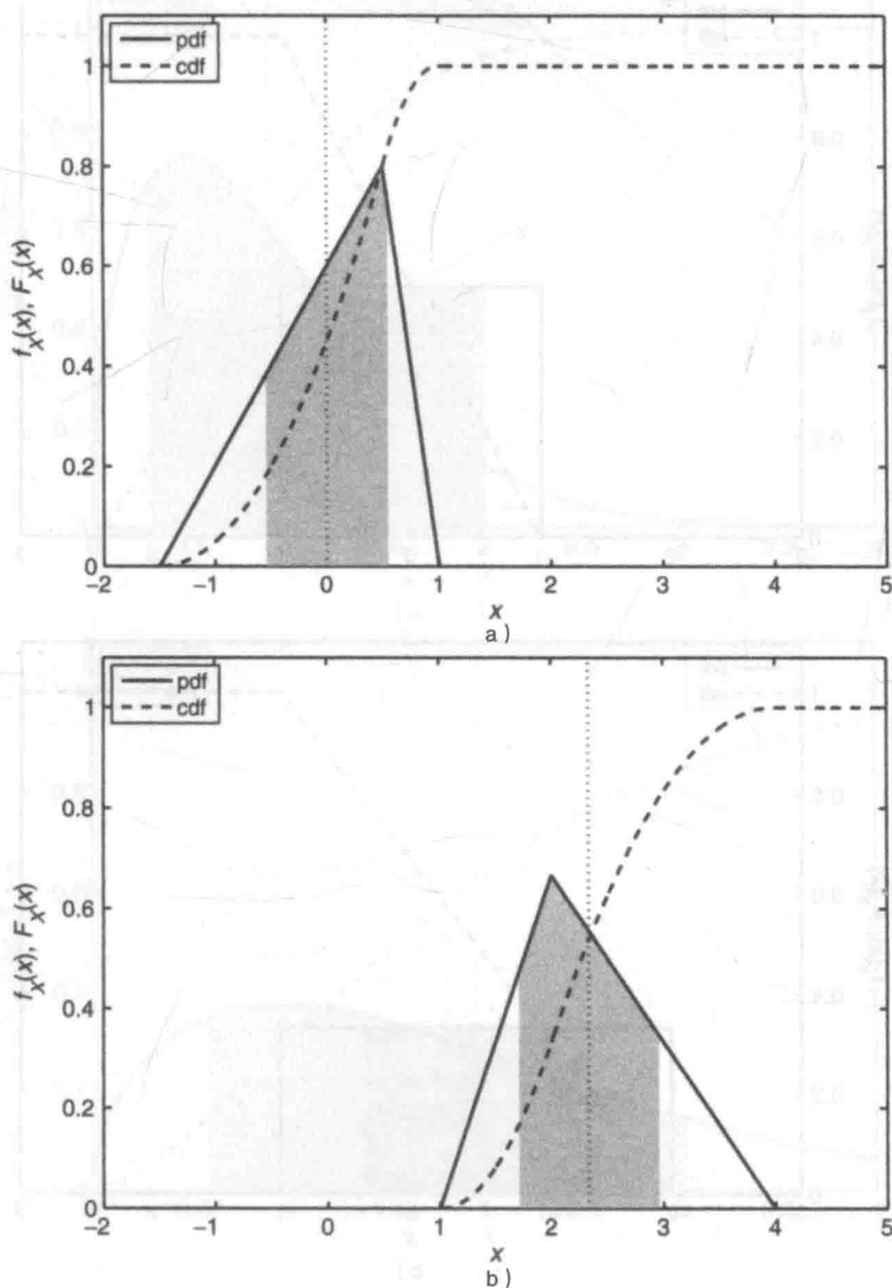


图 A-20 三角 pdf 和 cdf。a) 三角随机变量 $a=-1.5$，$b=1$，$c=0.5$；b) 三角随机变量 $a=1$，$b=4$，$c=2$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A.3.21　均匀(连续)

$$\text{参数 } a,b \in \mathcal{R}(\text{其中 } a<b),\text{定义域 } x\in[a,b]$$

$$f_X(x)=[1/(b-a)][u(x-a)-u(x-b)]$$

$$F_X(x)=[(x-a)/(b-a)]I_{[a,b]}(x)+I_{(b,\infty)}(x)$$

$$\mu_X=(a+b)/2,\quad \sigma_X^2=(b-a)^2/12$$

$$m_e=(a+b)/2,\quad m_o\in[a,b](\text{不唯一})$$

$$\mu_n=[b^{n+1}-a^{n+1}]/[(n+1)(b-a)],\quad \mu_{c,n}=\begin{cases}0, & n\text{ 奇数}\\ (b-a)^n/[(n+1)2^n], & n\text{ 偶数}\end{cases}$$

$$\gamma_1=0,\quad \gamma_2=-6/5=-1.2$$

$$\Phi_X(\omega)=[\exp(j\omega b)-\exp(j\omega a)]/[j\omega(b-a)],h(X)=\log(b-a)$$

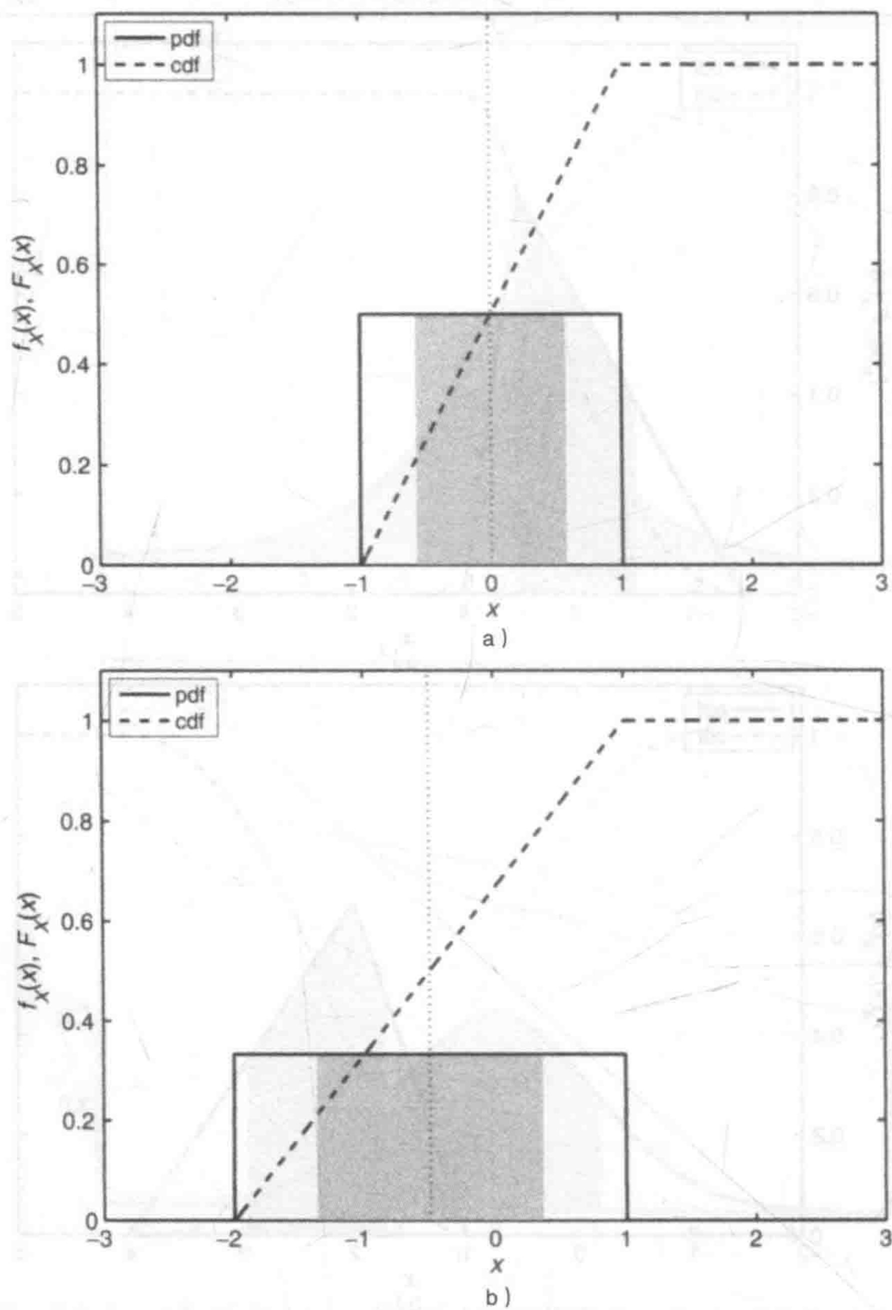


图 A-21　均匀(连续)pdf 和 cdf。a) 均匀(连续)随机变量 $a=-1$，$b=1$；b) 均匀(连续)随机变量 $a=-2$，$b=1$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A. 3. 22 韦伯

$$\text{参数 } r,\ \alpha>0,\ \text{定义域 } x\in\mathcal{R}^{+}$$

$$f_X(x)=(r/\alpha)(x/\alpha)^{r-1}\exp(-(x/\alpha)^r)u(x)$$

$$F_X(x)=[1-\exp(-(x/\alpha)^r)]u(x)$$

$$\mu_X=\alpha\Gamma(1+1/r),\quad \sigma_X^2=\alpha^2[\Gamma(1+2/r)-\Gamma^2(1+1/r)]$$

$$m_e=\alpha[\ln(2)]^{1/r},\quad m_o=\alpha((r-1)/r)^{1/r}(\text{其中 } r>1),\quad \mu_n=\alpha^n\Gamma(1+n/r)$$

$$\gamma_1=\frac{\Gamma(1+3/r)-3\Gamma(1+1/r)\Gamma(1+2/r)+2\Gamma^3(1+1/r)}{[\Gamma(1+2/r)-\Gamma^2(1+1/r)]^{3/2}}$$

$$\gamma_2=\frac{\Gamma(1+4/r)-4\Gamma(1+1/r)\Gamma(1+3/r)-3\Gamma^2(1+2/r)+12\Gamma^2(1+1/r)\Gamma(1+2/r)-6\Gamma^4(1+1/r)}{[\Gamma(1+2/r)-\Gamma^2(1+1/r)]^2}$$

$$\Phi_X(\omega)=\sum_{n=0}^{\infty}[(\mathrm{j}\omega\alpha)^n/n!]\Gamma(1+n/r),h(X)=1+(r-1)\gamma/r+\ln(\alpha/r)$$

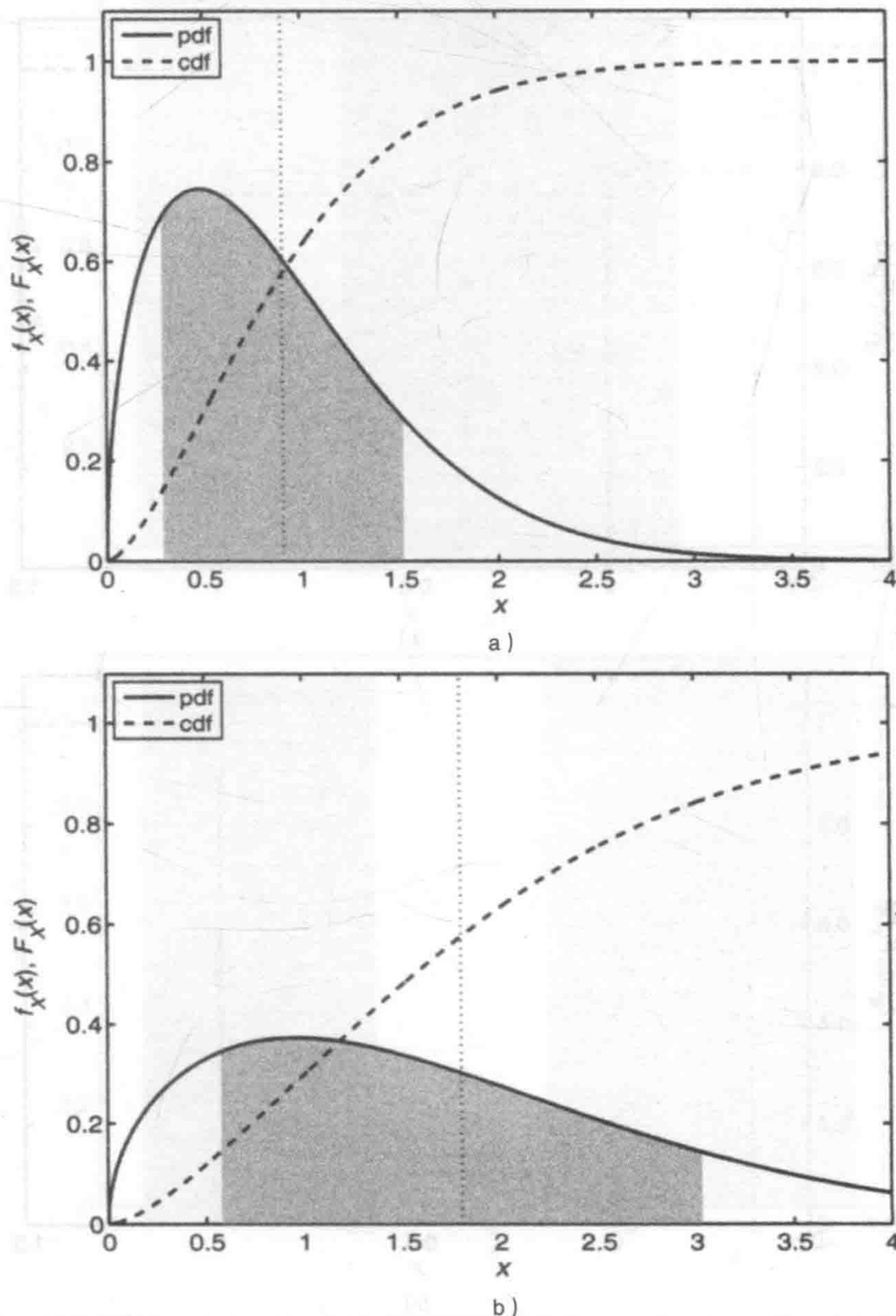


图 A-22 韦伯 pdf 和 cdf。a) 韦伯随机变量 $\alpha=1$，$r=1.5$；b) 韦伯随机变量 $\alpha=2$，$r=1.5$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A. 4. 1 伯努利(定义域为 {0,1})

参数 $p \in [0,1], q \triangleq 1-p$,定义域 $x \in \{0,1\}$

$$f_X(x) = q\delta(x) + p\delta(x-1), p_X[x] = p^x q^{1-x} I_{\{0,1\}}[x]$$

$$F_X(x) = qu(x) + pu(x-1)$$

$$\mu_X = p, \quad \sigma_X^2 = pq, \quad \mu_n = p, \quad \mu_{c,3} = pq(1-2p), \quad \mu_{c,4} = pq(1-3pq)$$

$$m_e = \text{不存在}, \quad m_o = \begin{cases} 0, & p \leqslant q \\ 1, & p \geqslant q \end{cases}$$

$$\gamma_1 = (1-2p)/\sqrt{pq}, \quad \gamma_2 = 1/pq - 6$$

$$\Phi_X(j\omega) = q + p\exp(j\omega), \quad H(X) = -p\log(p) - q\log(q)$$

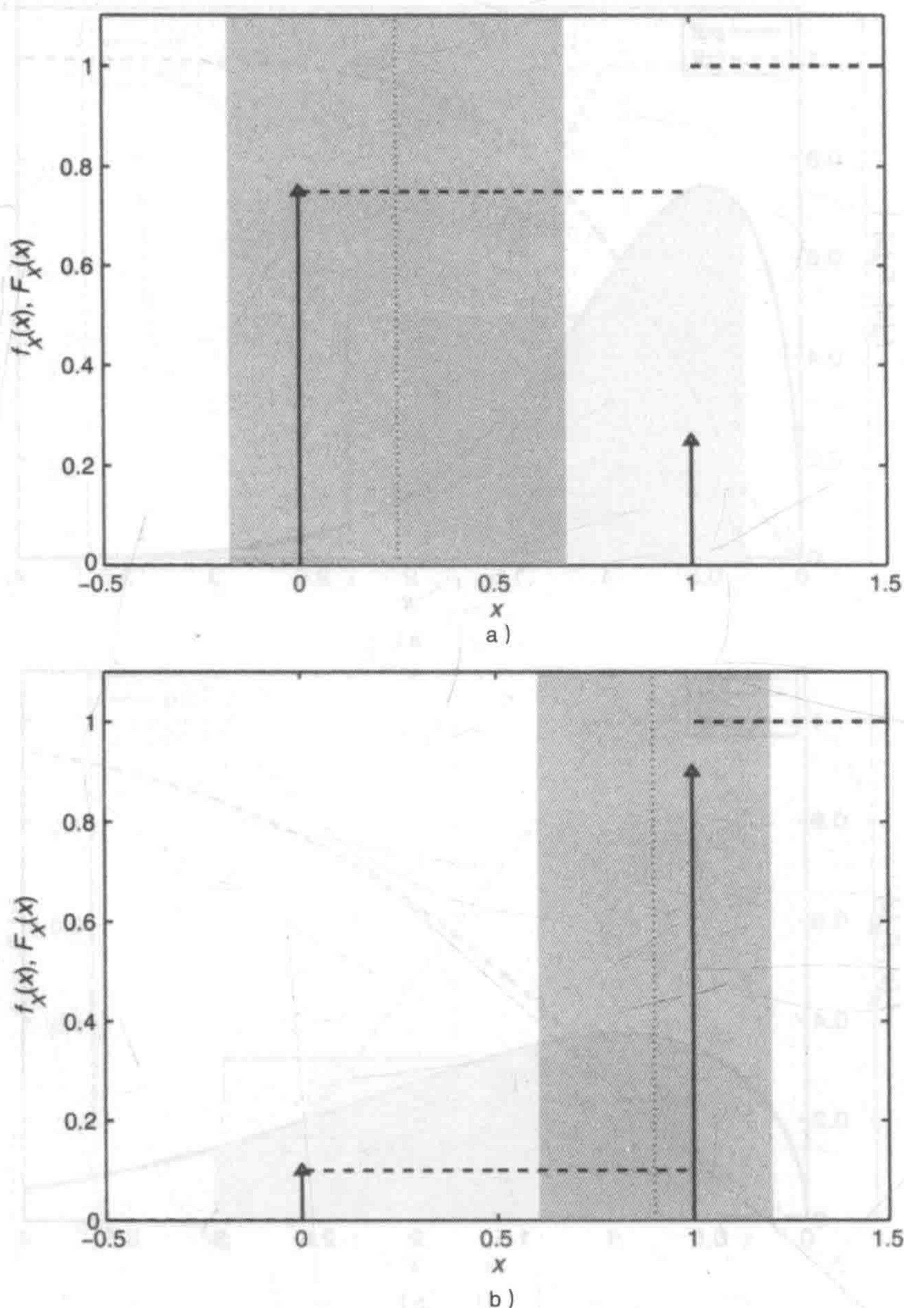


图 A-23 伯努利 pdf 和 cdf。a) 伯努利随机变量 $p=0.25$;b) 伯努利随机变量 $p=0.9$。虚线表示 μ_X,阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A.4.2 伯努利(对称，定义域为{-1，1})

$$参数\ p\in[0,1],\quad q\triangleq 1-p,\quad 定义域\ x\in\{-1,1\}$$

$$f_X(x)=q\delta(x+1)+p\delta(x-1),\quad p_X[x]=p^{(1+x)/2}q^{(1-x)/2}I_{\{-1,1\}}[x]$$

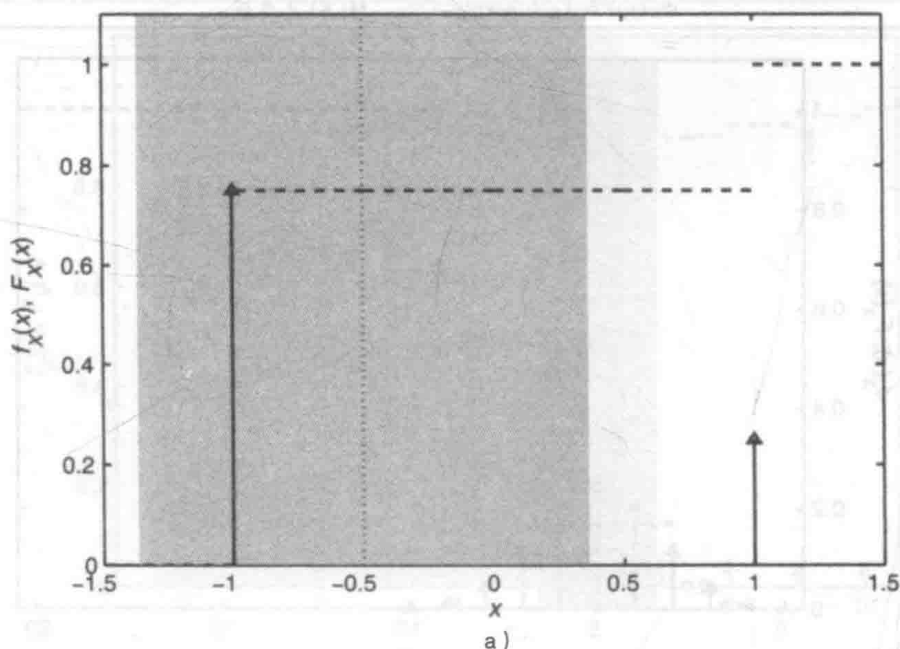
$$F_X(x)=qu(x+1)+pu(x-1),\quad \mu_X=p-q,\quad \sigma_X^2=1-(p-q)^2$$

$$\mu_{c,3}=2(p-q)[(p-q)^2-1],\quad \mu_{c,4}=1+2(p-q)^2-3(p-q)^4$$

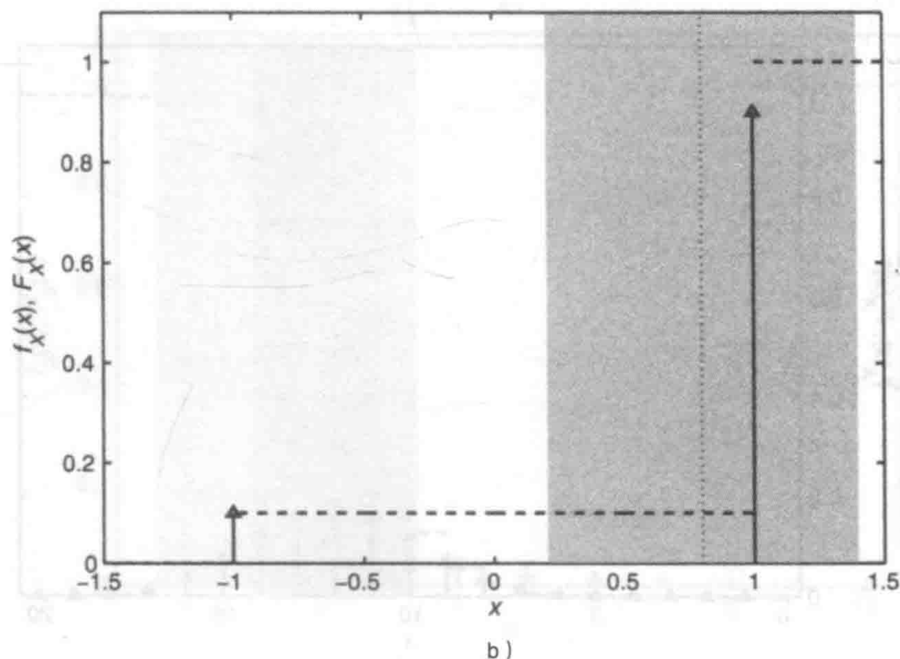
$$\mu_n=\begin{cases}p-q, & n\ 奇数\\ 1, & n\ 偶数\end{cases},\quad m_e\ 不存在,\quad m_o=\begin{cases}-1, & p\leqslant q\\ 1, & p\geqslant q\end{cases}$$

$$\gamma_1=-2(p-q)/\sqrt{1-(p-q)^2},\quad \gamma_2=-2[1-4(p-q)^2+3(p-q)^4]/[1-2(p-q)^2+(p-q)^4]$$

$$\Phi_X(j\omega)=q\exp(-j\omega)+p\exp(j\omega),H(X)=-p\log(p)-q\log(q)$$



a)



b)

图 A-24 对称伯努利 pdf 和 cdf。a) 对称伯努利随机变量 $p=0.25$；b) 对称伯努利随机变量 $p=0.9$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A.4.3 二项式

参数 $p\in[0, 1]$，$q\triangleq 1-p$，$N\in\mathcal{N}$，定义域 $x\in\{0, \cdots, N\}$

$$f_X(x)=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}\delta(x-n),\quad p_X[x]=\binom{N}{x}p^x q^{N-x}I_{\{0,\cdots,N\}}[x]$$

$$F_X(x)=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}u(x-n)=\sum_{n=0}^{\lfloor x\rfloor}\binom{N}{n}p^n q^{N-n}$$

$\mu_X=Np$，$\sigma_X^2=Npq$，$\mu_2=Np(q+Np)$，$m_e=\lfloor Np\rfloor$ 或者 $\lceil Np\rceil$

$$m_o=\begin{cases}\lfloor (N+1)p\rfloor, & \text{非整数 } (N+1)p\\ (N+1)p \text{ 和 } (N+1)p-1, & \text{整数 } (N+1)p\end{cases}$$

$\mu_{c,3}=Npq(q-p)$，$\gamma_1=(1-2p)/\sqrt{Npq}$，$\gamma_2=(1-6pq)/Npq$

$\Phi_X(j\omega)=[q+p\exp(j\omega)]^N$，$H(X)$ 太复杂

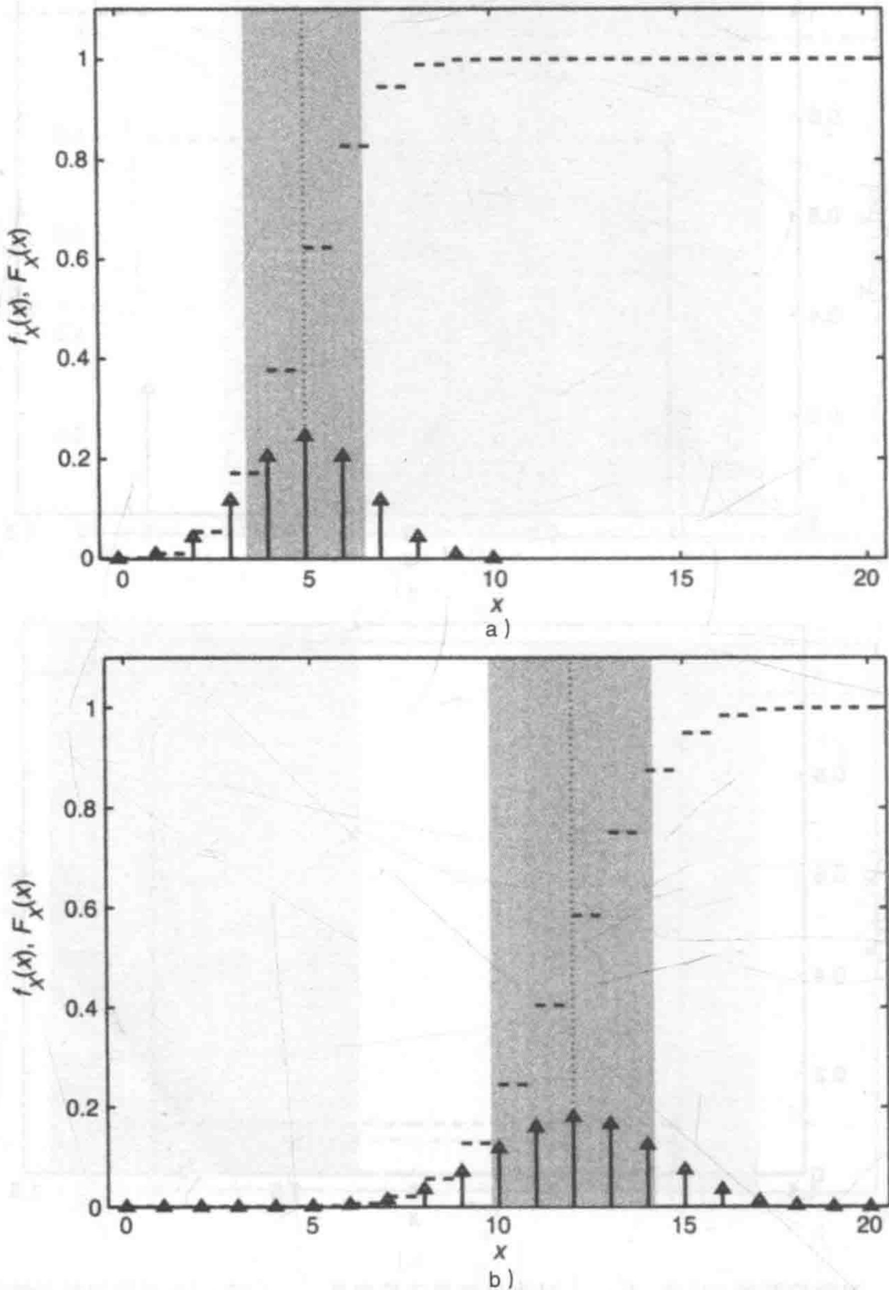


图 A-25 二项式 pdf 和 cdf。a) 二项式随机变量 $N=10$，$p=0.5$；b) 二项式随机变量 $N=20$，$p=0.6$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.4.4 几何(定义域$\mathcal{Z}^+$)

参数 $p\in[0,1]$，$q\triangleq 1-p$，定义域 $x\in\mathcal{Z}^+$

$$f_X(x)=\sum_{n=0}^{\infty}pq^n\delta(x-n),p_X[x]=pq^xI_{\mathcal{Z}^+}[x]$$

$$F_X(x)=\sum_{n=0}^{\infty}pq^nu(x-n)=(1-q^{x+1})I_{\mathcal{Z}^+}[x]$$

$$\mu_X=q/p,\quad \sigma_X^2=q/p^2,m_e=\lceil -1/\log_2(q)\rceil-1,m_o=0$$

$$\mu_2=q(2-p)/p^2,\quad \mu_3=q(6q+p^2)/p^3,\quad \mu_{c,3}=q(2-p)/p^3,\quad \mu_{c,4}=q(9q+p^2)/p^4$$

$$\gamma_1=(2-p)/\sqrt{q},\quad \gamma_2=p^2/q+6$$

$$\Phi_X(\mathrm{j}\omega)=p[1-q\exp(\mathrm{j}\omega)]^{-1},\quad H(X)=-\log(p)-(q/p)\log(q)$$

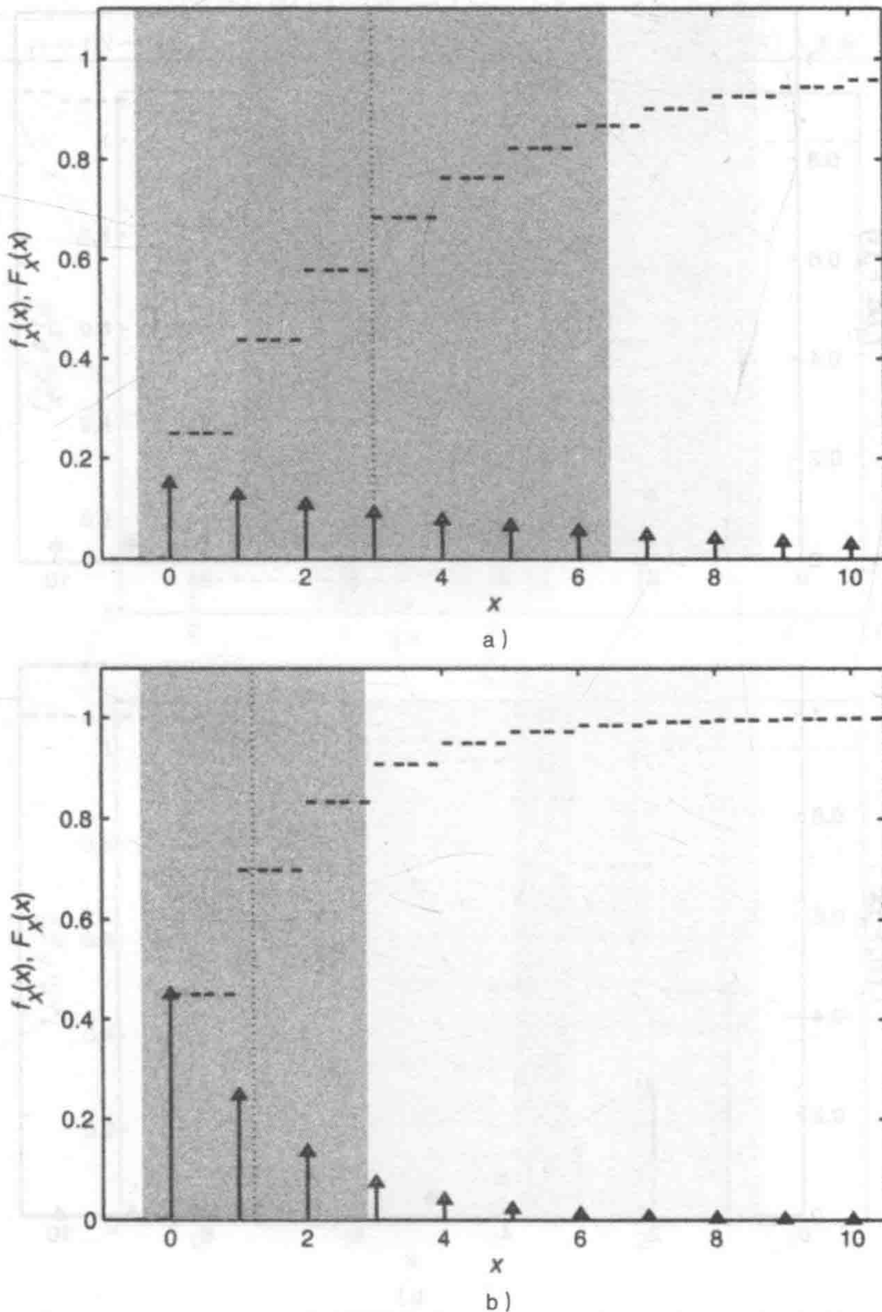


图 A-26 几何 pdf 和 cdf。a) 几何随机变量 $p=0.25$；b) 几何随机变量 $p=0.45$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A. 4. 5 几何(定义域变为$\mathcal{N}$)

参数 $p\in[0,1]$，$q\triangleq 1-p$，定义域 $x\in\mathcal{N}$

$$f_X(x)=\sum_{n=1}^{\infty}pq^{n-1}\delta(x-n),p_X[x]=pq^{x-1}I_{\mathcal{N}}[x]$$

$$F_X(x)=\sum_{n=1}^{\infty}pq^{n-1}u(x-n)=(1-q^x)I_{\mathcal{N}}[x]$$

$$\mu_X=1/p,\quad \sigma_X^2=q/p^2,\quad m_e=\lceil -1/\log_2(q)\rceil,\quad m_o=1$$

$$\mu_2=(2-p)/p^2,\quad \mu_3=(6q+p^2)/p^3,\quad \mu_{c,3}=q(2-p)/p^3,\quad \mu_{c,4}=q(9q+p^2)/p^4$$

$$\gamma_1=(2-p)/\sqrt{q},\quad \gamma_2=p^2/q+6$$

$$\Phi_X(j\omega)=p\exp(j\omega)[1-q\exp(j\omega)]^{-1},\quad H(X)=-\log(p)-(q/p)\log(q)$$

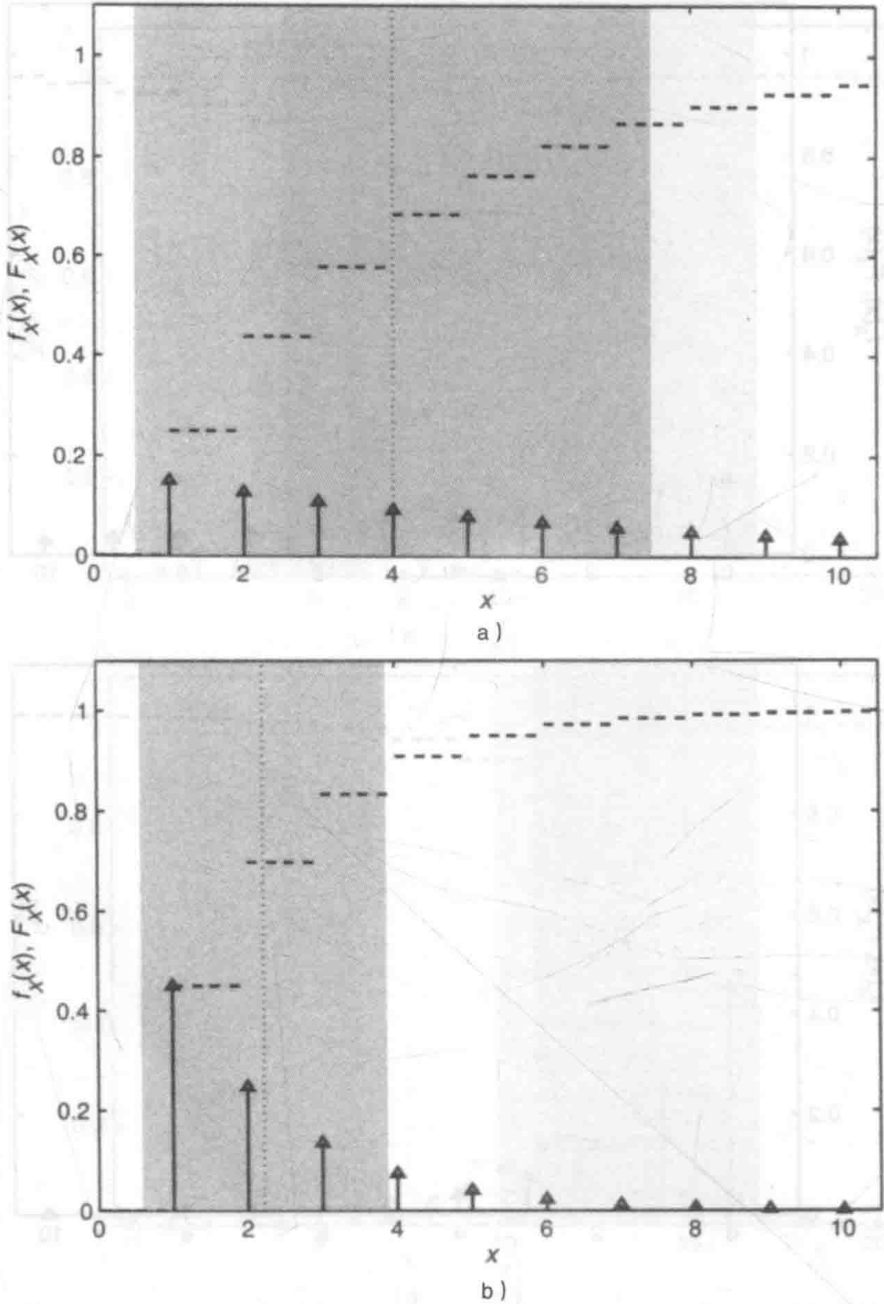


图 A-27 移位几何 pdf 和 cdf。a) 移位几何随机变量 $p=0.25$；b) 移位几何随机变量 $p=0.45$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A.4.6　超几何

参数 $N \in \mathcal{N}$，$M \in \{0,\cdots,N\}$，$n \in \{1,\cdots,N\}$，

定义域 $x \in \{x_L \triangleq \max(0, n+M-N), \cdots, \min(M,n) \triangleq x_H\}$

$$f_X(x) = \sum_{m=x_L}^{x_H} \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \delta(x-m), \quad p_X[x] = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} I_{\{x_L,\cdots,x_H\}}[x]$$

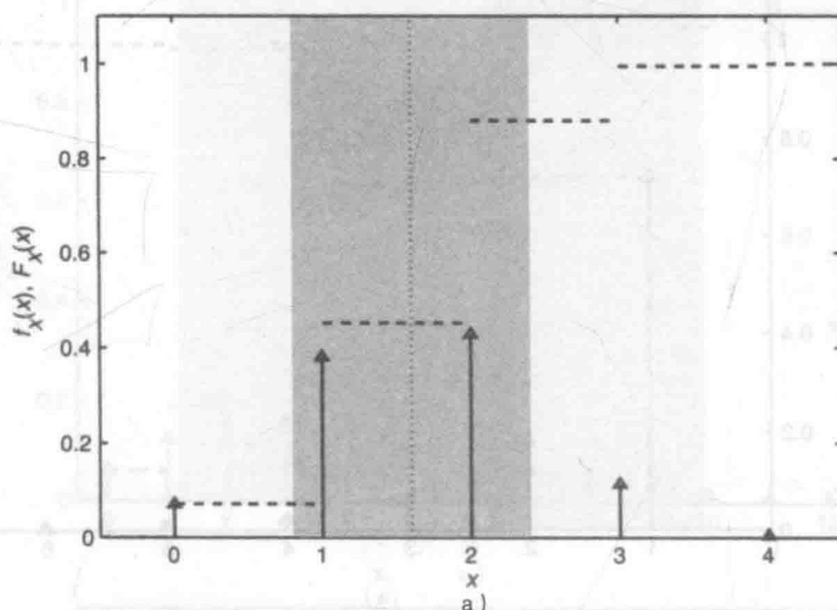
$$F_X(x) = \sum_{m=x_L}^{x_H} \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} u(x-m) = \sum_{m=x_L}^{\lfloor x \rfloor} \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

$\mu_X = nM/N$，$\sigma_X^2 = nM(N-M)(N-n)/N^2(N-1)$，$m_e$ 隐式

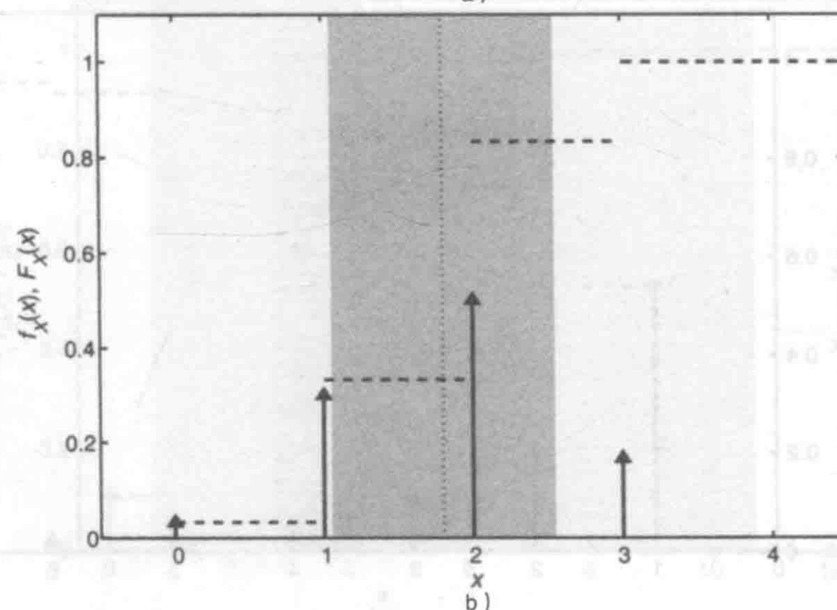
$\mu_{c,3} = nM(N-M)(N-2M)(N-n)(N-2n)/N^3(N-1)(N-2)$

$m_o = \lfloor (n+1)(M+1)/(N+2) \rfloor$，$\gamma_2$ 太复杂，$\Phi_X(j\omega)$ 太复杂

$\gamma_1 = (N-2M)(N-2n)\sqrt{N-1}/(N-2)\sqrt{nM(N-M)(N-n)}$，$H(X)$ 太复杂



a)



b)

图 A-28　超几何 pdf 和 cdf。a) 超几何随机变量 $N=10$，$M=n=4$；b) 超几何随机变量 $N=10$，$M=6$，$n=3$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.4.7 对数(Log 级数)

参数 $p\in[0,1]$，$q\triangleq 1-p$，定义域 $x\in\mathcal{N}$

$$f_X(x)=[-1/\ln(q)]\sum_{N=1}^{\infty}(p^n/n)\delta(x-n),\quad p_X[x]=[-p^x/x\ln(q)]I_{\mathcal{N}}[x]$$

$$F_X(x)=[-1/\ln(q)]\sum_{N=1}^{\infty}(p^n/n)u(x-n)=[-1/\ln(q)]\sum_{n=1}^{\lfloor x\rfloor}p^n/n$$

$$\mu_X=-p/q\ln(q),\quad \sigma_X^2=-p[p+\ln(q)]/q^2\ln^2(q)$$

$$m_e\text{ 为隐式},\quad m_o=1,\quad \mu_2=-p/q^2\ln(q),\quad \mu_3=-p(p+1)/q^3\ln(q)$$

$$\gamma_1=[2p^2+3p\ln(q)+(p+1)\ln^2(q)]/[p+\ln(q)]\sqrt{-p\ln(q)-p^2}$$

$$\gamma_2=-[3p^3+6p^2\ln(q)+4p(p+1)\ln^2(q)+(p^2+4p+1)\ln^3(q)]/p[p+\ln(q)]^2$$

$$\Phi_X(j\omega)=\ln(1-p\exp(j\omega))/\ln(q)$$

$$H(X)=(p/q)\ln(p/q)-\ln(-1/\ln(q))-[1/\ln(q)]\sum_{N=1}^{\infty}\ln(n)p^n/n$$

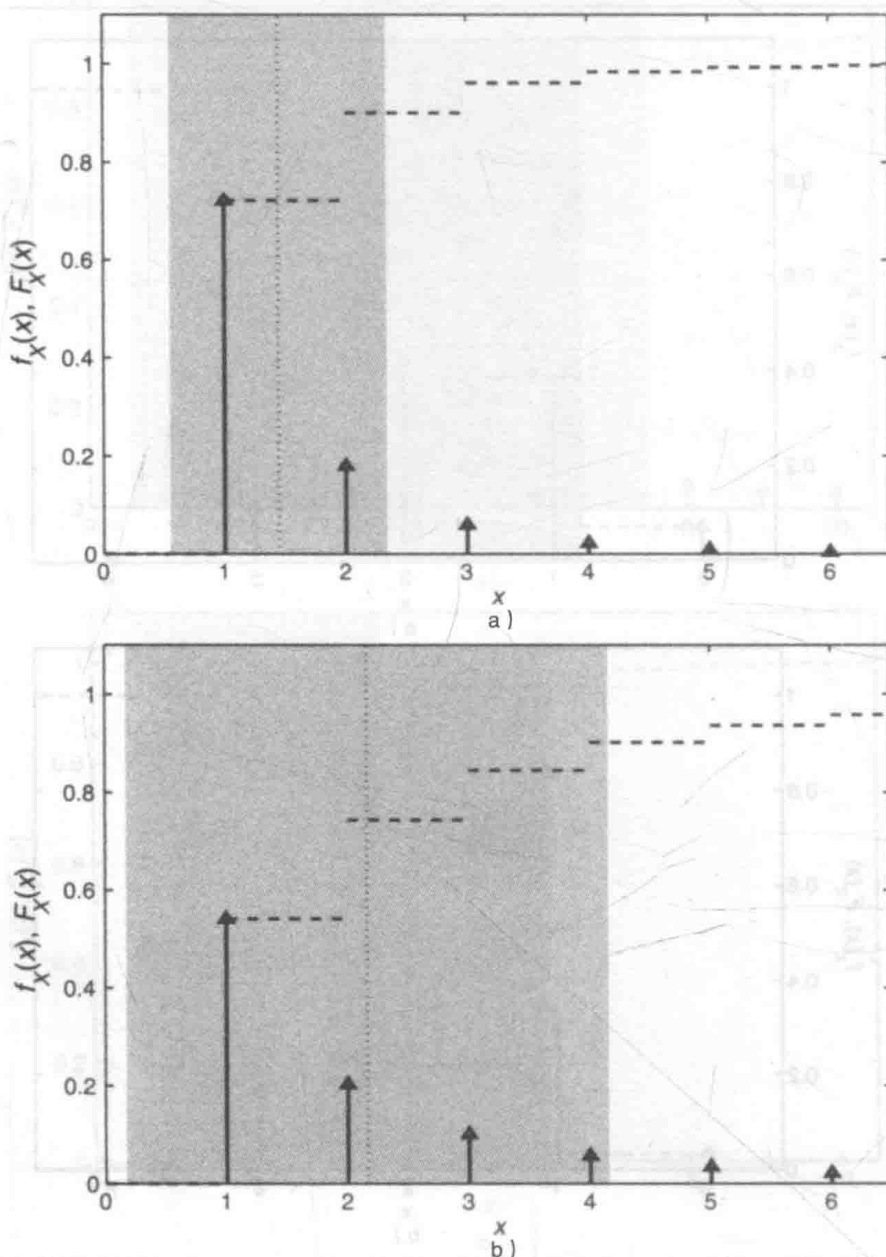


图 A-29 对数 pdf 和 cdf。a) 对数随机变量 $p=0.5$；b) 对数随机变量 $p=0.75$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

A.4.8 负二项式(帕斯卡)

参数 $p \in [0, 1]$, $q \triangleq 1-p$, $N \in \mathcal{N}$。定义域 $x \in \mathcal{Z}^+$

$$f_X(x) = \sum_{n=0}^{\infty} \binom{n+N-1}{n} p^N q^n \delta(x-n), p_X[x] = \binom{x+N-1}{x} p^N q^x I_{\mathcal{Z}^+}[x]$$

$$F_X(X) = \sum_{n=0}^{\infty} \binom{n+N-1}{n} p^N q^n u(x-n) = \sum_{n=0}^{\lfloor x \rfloor} \binom{n+N-1}{n} p^N q^n$$

$$\mu_X = Nq/p, \quad \sigma_X^2 = Nq/p^2$$

m_e 为隐式, $m_o = 0$ 或 $\lfloor N-1)q/p \rfloor$(其中 $N > 1$), $\mu_{c,3} = Nq(1+q)/p^3$

$$\mu_{c,4} = 3N^2 q^2/p^4 + Nq(p^2+6q)/p^4, \quad \gamma_1 = (2-p)/\sqrt{Nq}, \quad \gamma_2 = p^2/Nq + 6/N$$

$\Phi_X(j\omega) = p^N/[1-q\exp(j\omega)]^N$, $H(X)$ 太复杂

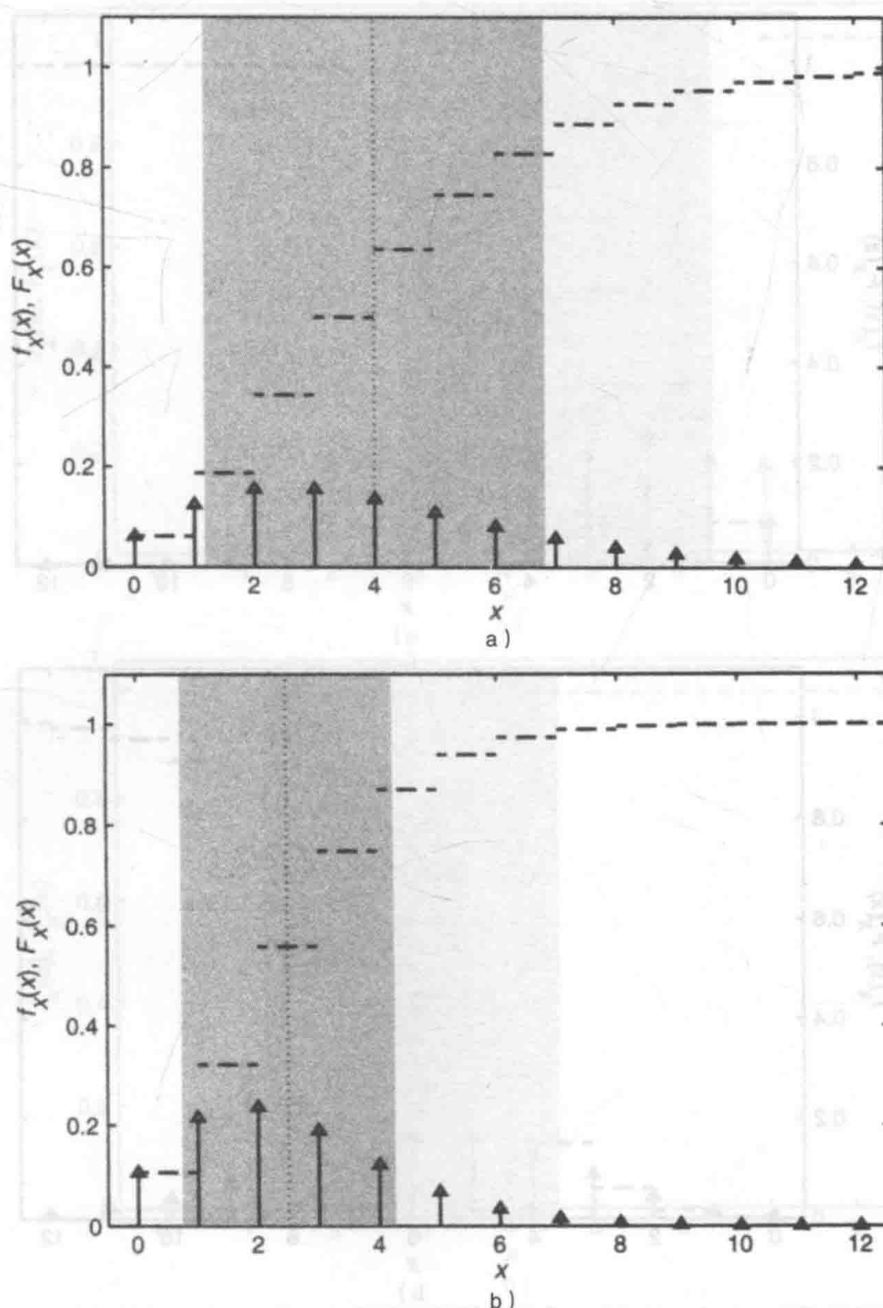


图 A-30 负二项式 pdf 和 cdf。a) 负二项式随机变量 $N=4$, $p=0.5$; b) 负二项式随机变量 $N=10$, $p=0.8$。虚线表示 μ_X, 阴影部分表示 $[\mu_X - \sigma_X, \mu_X + \sigma_X]$

A.4.9 泊松

参数 $\alpha>0$，定义域 $x\in\mathcal{Z}^+$

$$f_X(x)=\sum_{n=0}^{\infty}\exp(-\alpha)(\alpha^n/n!)\delta(x-n),\quad p_X[x]=\exp(-\alpha)(\alpha^x/x!)I_{\mathcal{Z}^+}[x]$$

$$F_X(x)=\sum_{n=0}^{\infty}\exp(-\alpha)(\alpha^n/n!)u(x-n)=\sum_{n=0}^{\lfloor x\rfloor}\exp(-\alpha)(\alpha^n/n!)$$

$$\mu_X=\alpha,\quad \sigma_X^2=\alpha,\quad m_e\text{ 为隐式},\quad m_o=\begin{cases}\alpha\text{ 和 }\alpha-1, & \alpha\in\mathcal{N}\\ \lfloor\alpha\rfloor, & \text{其他}\end{cases}$$

$$\mu_{c,3}=\alpha,\quad \mu_{c,4}=3\alpha^2+\alpha,\quad \mu_{c,5}=10\alpha^2+\alpha,\quad \gamma_1=1/\sqrt{\alpha},\quad \gamma_2=1/\alpha$$

$$\Phi_X(\mathrm{j}\omega)=\exp(\alpha[\exp(\mathrm{j}\omega)-1]),\quad H(X)=\exp(-\alpha)\sum_{n=0}^{\infty}\alpha^n\log(n!)/n!+\alpha[1-\log(\alpha)]$$

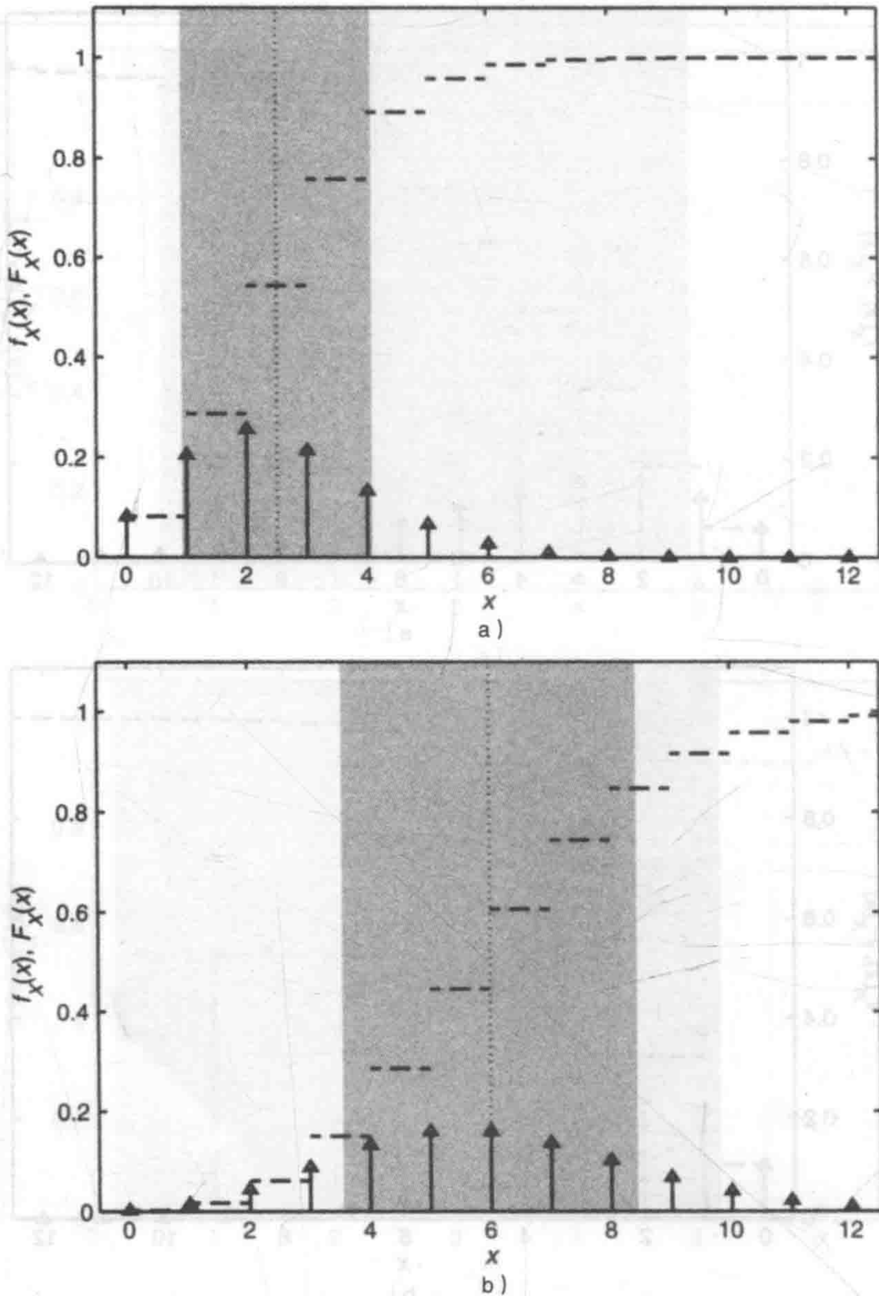


图 A-31 泊松 pdf 和 cdf。a) 泊松随机变量 $\alpha=2.5$；b) 泊松随机变量 $\alpha=6$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\ \mu_X+\sigma_X]$

A.4.10 均匀(离散)

参数 $a,b \in \mathbb{Z}$(其中 $a<b$)， $N \triangleq b-a+1$， 定义域 $x \in \{a,\cdots,b\}$

$$f_X(x) = (1/N)\sum_{n=a}^{b}\delta(x-n), \quad p_X[x] = (1/N)I_{\{a,\cdots,b\}}[x]$$

$$F_X(x) = (1/N)\sum_{n=a}^{b}u(x-n) = (1/N)(\lfloor x \rfloor - a)I_{[a,b]}(x) + I_{(b,\infty)}(x)$$

$$\mu_X = (a+b)/2, \quad \sigma_X^2 = (N^2-1)/12, \quad m_e = \begin{cases} \lfloor (a+b+1)/2 \rfloor, & (a+b)\text{偶数}(N\text{奇数}) \\ \text{不存在}, & (a+b)\text{奇数}(N\text{偶数}) \end{cases}$$

$$m_o \in \{a,\cdots,b\}(\text{不唯一}), \quad \mu_{c,3} = 0, \quad \mu_{c,4} = (N^2-1)(3N^2-7)/240$$

$$\gamma_1 = 0, \quad \gamma_2 = -(6/5)(N^2+1)/(N^2-1)$$

$$\Phi_X(j\omega) = (1/N)\exp(j\omega a)[1-\exp(j\omega N)]/[1-\exp(j\omega)], \quad H(X) = \log(N)$$

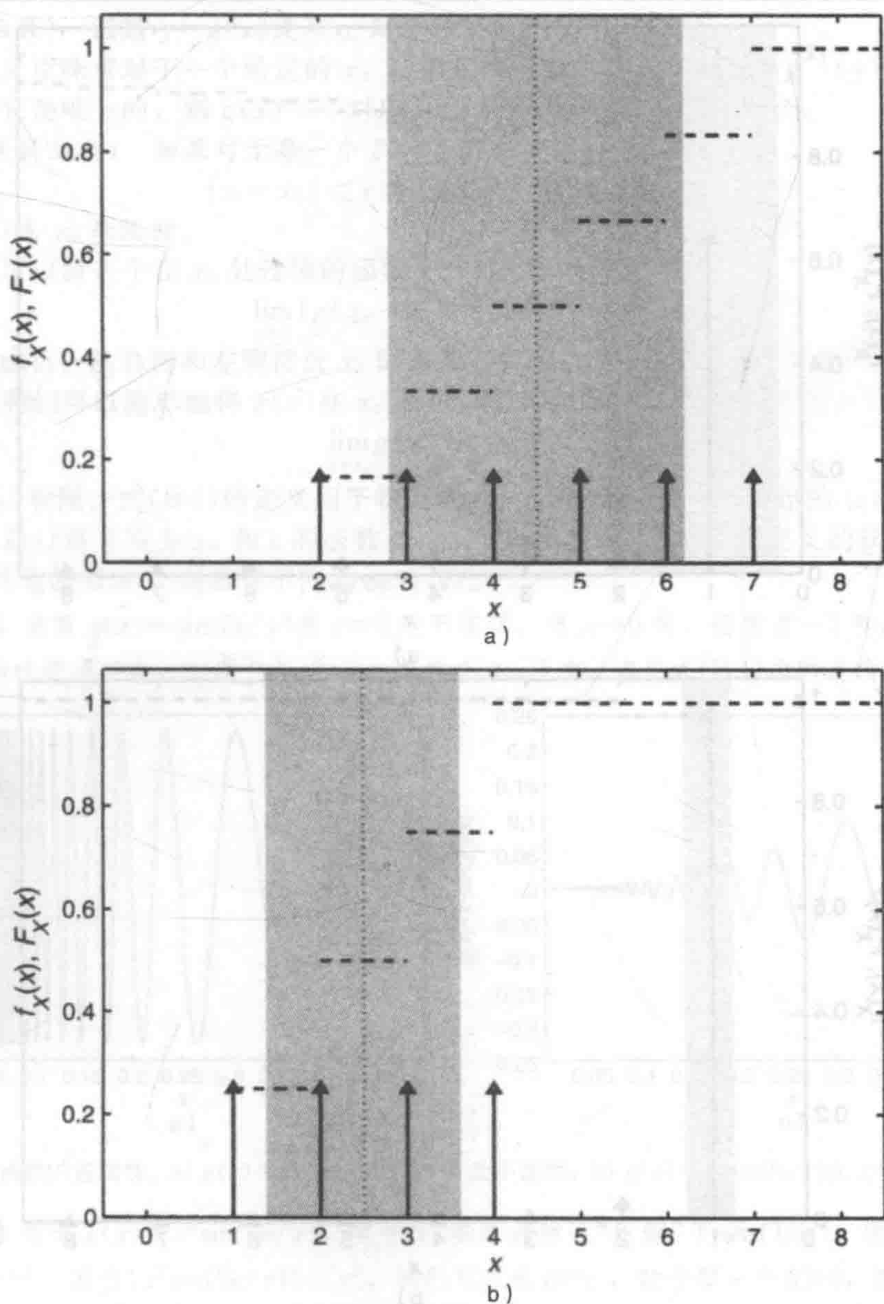


图 A-32 离散均匀 pdf 和 cdf。a) 离散均匀随机变量 $a=2$，$b=7$；b) 离散均匀随机变量 $a=1$，$b=4$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X, \mu_X+\sigma_X]$

A.4.11　泽塔(Zipf)

参数 $\alpha\in(1,\infty)$，定义域 $x\in\mathcal{N}$

$$f_X(x)=[1/\zeta(\alpha)]\sum_{n=1}^{\infty}(1/n^{\alpha})\delta(x-n),p_X[x]=[1/\zeta(\alpha)x^{\alpha}]I_{\mathcal{N}}[x]$$

$$F_X(x)=[1/\zeta(\alpha)]\sum_{n=1}^{\infty}(1/n^{\alpha})u(x-n)=[1/\zeta(\alpha)]\sum_{n=1}^{\lfloor x\rfloor}1/n^{\alpha}$$

$\mu_X=\zeta(\alpha-1)/\zeta(\alpha)$（其中 $\alpha>2$），$\sigma_X^2=[\zeta(\alpha)\zeta(\alpha-2)-\zeta^2(\alpha-1)]/\zeta^2(\alpha)$（其中 $\alpha>3$）

m_e 为隐式，$m_o=1$，$\mu_n=\zeta(\alpha-n)/\zeta(\alpha)$（其中 $\alpha>n+1$）

$\gamma_1=$ 太复杂，$\gamma_2=$ 太复杂，$\Phi_X(j\omega)=[1/\zeta(\alpha)]\sum_{n=1}^{\infty}\exp(j\omega n)/n^{\alpha}$

$$H(X)=[1/\zeta(\alpha)]\sum_{n=1}^{\infty}\log(n^{\alpha}\zeta(\alpha))/n^{\alpha}$$

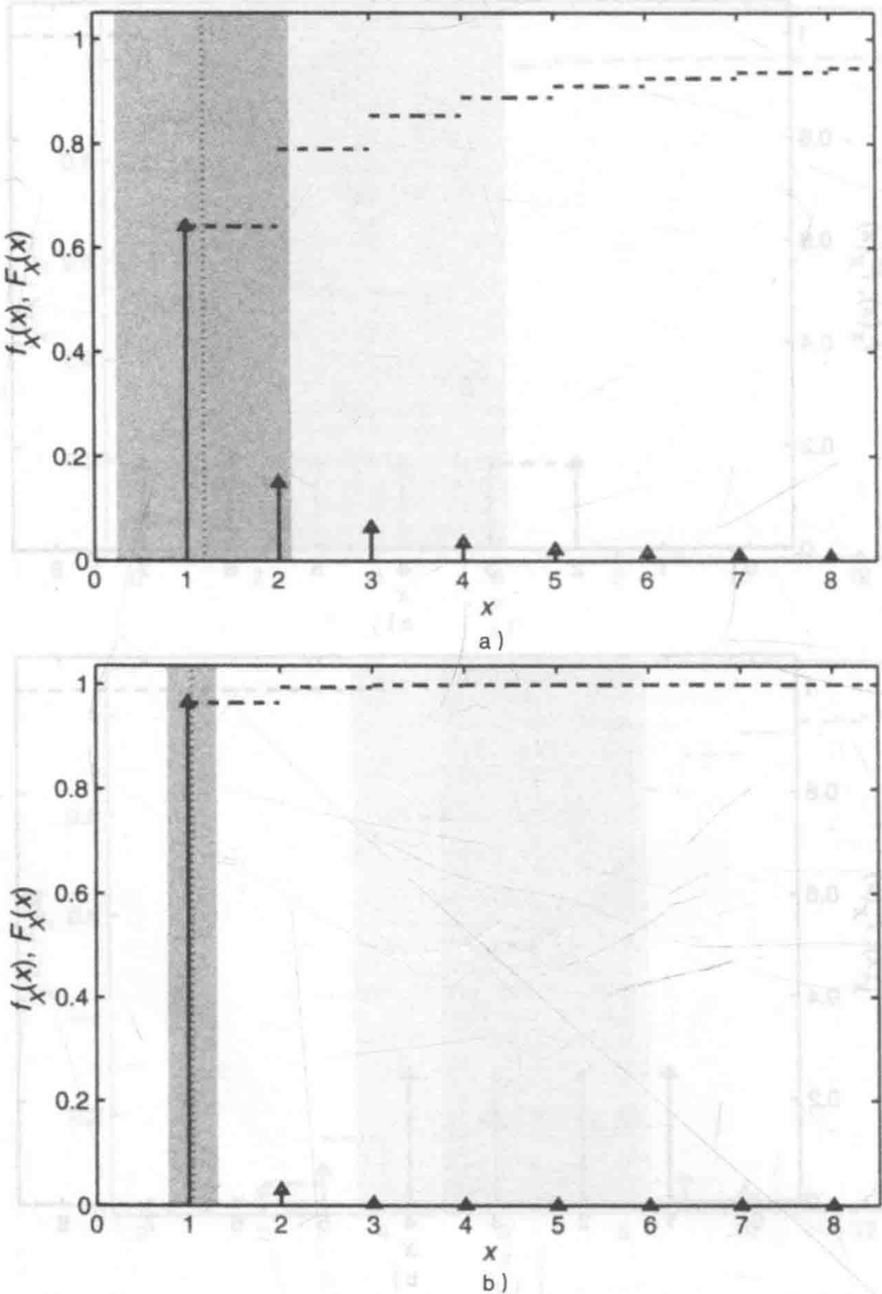


图 A-33　泽塔 pdf 和 cdf。a) 泽塔随机变量 $\alpha=3.5$；b) 泽塔随机变量 $\alpha=5$。虚线表示 μ_X，阴影部分表示 $[\mu_X-\sigma_X,\mu_X+\sigma_X]$

附录 B 函数和属性

本附录中，我们回顾了函数的一些基本属性，然后对书中几个章节用到的特殊函数以及符号进行了总结。

B.1 连续性与有界变分

定义(函数) 函数 $y=g(x)$是从 x 到 y 的映射。

这个定义意味着对于一个给定的 x，映射到唯一的 y。逆变换 $x=g^{-1}(y)$可能不是唯一的；如果它是唯一的，则 $g(x)$一一对应。

定义(连续函数) 如果对于每一个 $\delta>0$，存在 $\varepsilon>0$，使得

$$|x-x_0|<\varepsilon \Rightarrow |g(x)-g(x_0)|<\delta \tag{B-1}$$

则函数 $g(x)$在 x_0 处连续。

我们也可以将一个在 x_0 处连续的函数 $g(x)$写为

$$\lim_{\varepsilon\to 0}|g(x_0+\varepsilon)-g(x_0)|=0 \tag{B-2}$$

其中 ε 可正或负：从右侧和左侧接近 x_0 时极限必须成立。

因此，我们可以简单地将 $g(x)$在 x_0 处连续写为：

$$\lim_{x\to x_0} g(x)=g(x_0) \tag{B-3}$$

上式中 $g(x_0)$有限。式(B-1)的定义用于特定的点 x_0，意味着将一个特定的$\{\varepsilon,\delta\}$对用于该点。这可以通过将 δ 写为 x_0 和 ε 的函数 $\delta(x_0,\varepsilon)$ 来强调(不要与后面定义的狄拉克 δ 函数混淆)。对其他的点 x_1，则需要不同的 $\delta(x_1,\varepsilon)$。

例 B-1 函数 $g(x)=\sin(2\pi/x)$在 $x=0$ 处不连续，当 $x\to 0$ 时，函数在 -1 和 1 之间振荡，在 $x=0$ 处的极限不存在，如图 B-1a 所示。对每个 δ，没有 ε 满足式(B-1)中的连续性条件。◀

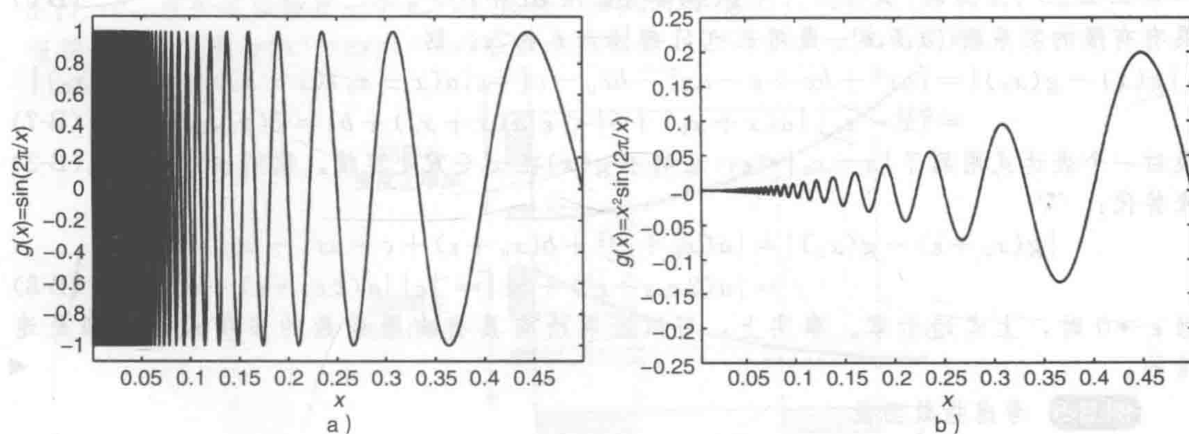


图 B-1 函数的连续性。a) $g(x)=\sin(2\pi/x)$在 $x=0$ 处不连续；b) $g(x)=x^2\sin(2\pi/x)$在 $x=0$ 处连续

例 B-2 考虑 $g(x)=x^2\sin(2\pi/x)$，其中 x^2 在原点附近“压缩”了 $\sin(1/x)$，因此 $g(0)=0$，如图 B-1b 所示。因为$|x^2\sin(2\pi/x)|\leqslant x^2$，我们可以选 $\delta=\varepsilon^2$，对于每一个 $\delta>0$，连续性条件变为：$|x-0|<\varepsilon\Rightarrow|x^2\sin(2\pi/x)-0|\leqslant x^2=\varepsilon^2$，则对应 $\delta(x=0,\varepsilon)$ ◀

如果式(B-1)只对正的或负的 ε 成立(而不是都成立)，则 $g(x)$分别为右连续或左连续。

定义(右或左连续函数) 如果对每个$\delta>0$，存在$\varepsilon>0$，使得

$$x_0<x<x_0+\varepsilon \Rightarrow |g(x)-g(x_0)|<\delta \tag{B-4}$$

则函数$g(x)$是右连续。

同样，一个左连续函数满足：

$$x_0-\varepsilon<x<x_0 \Rightarrow |g(x)-g(x_0)|<\delta \tag{B-5}$$

显然，一个连续函数在x_0处既是左连续又是右连续的。图 B-2 中显示的是一个右连续函数的例子。

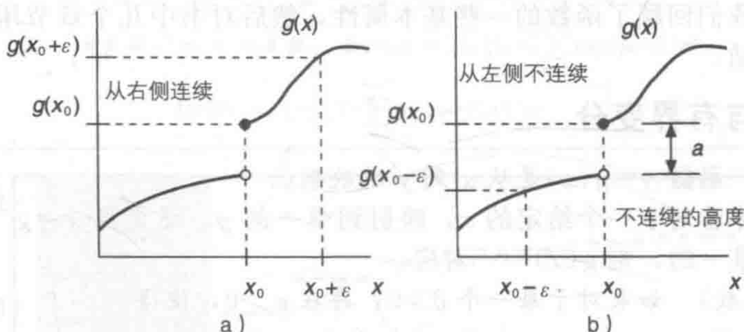


图 B-2 在x_0处右连续的函数。a) 实心圆圈定义了$g(x_0)$，使得当$\varepsilon\to 0$时，从右边有$g(x+\varepsilon)\to g(x_0)$；b) 从左侧，$g(x_0-\varepsilon)$趋近$g(x_0)-a$，其中$a>0$是不连续的高度。空心圆圈表示该函数不是左连续的：当$\varepsilon\to 0$时，$g(x_0-\varepsilon)$"跳变"到$g(x_0)$

例 B-3 离散随机变量的累积分布函数(cdf)被定义为右连续。图 A-25 给出了附录 A 中总结的二项式随机变量分布的例子。考虑 cdf 在$x=1$的不连续性，此处的概率质量为$\binom{N}{1}pq^{N-1}$。很显然，对每个$\delta>0$，我们可以找到$\varepsilon>0$，使得当$1<x<1+\varepsilon$时，$|F_X(x)-F_X(1)|<\delta$。实际上对$1<x<2$，$|F_X(x)-F_X(1)|=0$。而另一方面，找不到一个$\varepsilon>0$，使得$x=1-\varepsilon$无限接近1，并且对每个δ都满足$|F_X(1)-F_X(x)|<\delta$，因为在$x=1$处有"跳变"。◀

例 B-4 考虑二次函数

$$g(x)=ax^2+bx+c \tag{B-6}$$

具有有限的实系数$\{a,b,c\}$。设对$\varepsilon>0$，有$|x-x_0|<\varepsilon$。则

$$\begin{aligned}|g(x)-g(x_0)|&=|ax^2+bx+c-ax_0^2-bx_0-c|=|a(x-x_0)(x+x_0)+b(x-x_0)|\\&=|x-x_0||a(x+x_0)+b|<\varepsilon|a(x+x_0)+b|=\delta(x_0,\varepsilon)\end{aligned} \tag{B-7}$$

最后一个表达式用到了$|x-x_0|<\varepsilon$，证明了$g(x)$在$x_0\in\mathcal{R}$处连续。我们也可以用式(B-2)来替代：

$$\begin{aligned}|g(x_0+\varepsilon)-g(x_0)|&=|a(x_0+\varepsilon)^2+b(x_0+\varepsilon)+c-ax_0^2-bx_0-c|\\&=|a(2x_0\varepsilon+\varepsilon^2)+b\varepsilon|=|\varepsilon||a(2x_0+\varepsilon)+b|\end{aligned} \tag{B-8}$$

当$\varepsilon\to 0$时，上式趋于零。事实上，可以证明所有具有有限系数的多项式函数都是连续的。◀

例 B-5 考虑指数函数

$$g(x)=\exp(-\lambda x) \tag{B-9}$$

其中$\lambda>0$，

$$\begin{aligned}|g(x_0+\varepsilon)-g(x_0)|&=|\exp(-\lambda(x_0+\varepsilon))-\exp(-\lambda x_0)|\\&=\exp(-\lambda x_0)|\exp(-\lambda\varepsilon)-1|\end{aligned} \tag{B-10}$$

当$\varepsilon\to 0$时，这个表达式趋向于0。则$\exp(-\lambda x_0)$在$x_0\in\mathcal{R}$处连续。

定义(右极限和左极限) 函数$g(x)$在x_0的**右极限**为b

$$\lim_{x\to x_0^+} g(x) = b \tag{B-11}$$

左极限为 a

$$\lim_{x\to x_0^-} g(x) = a \tag{B-12}$$

对于一个在 x_0(例如右和左连续)处连续的函数，有 $a=b$。◀

例 B-6 底函数 $\lfloor x \rfloor$(稍后定义)是不超过 x 的最大的整数。对于整数 $x_0=n$，它有以下右和左极限：

$$\lim_{x\to n^+} \lfloor x \rfloor = n, \quad \lim_{x\to n^-} \lfloor x \rfloor = n-1 \tag{B-13}$$

它是一个右连续函数的例子。顶函数 $\lceil x \rceil$(也在稍后定义)是不小于 x 的最小整数。对于整数 $x_0=n$，它有以下右和左极限：

$$\lim_{x\to n^+} \lceil x \rceil = n+1, \quad \lim_{x\to n^-} \lceil x \rceil = n \tag{B-14}$$

这意味着它是左连续的。◀

更强类型的连续允许将相同的 δ 运用于闭区间 $[a,b]$ $(b>a)$上任意点。

定义(一致连续)　函数 $g(x)$在$[a, b]$上一致连续，如果对每个 $\delta>0$，存在 $\varepsilon>0$，使得

$$|x_1-x_2|<\varepsilon \Rightarrow |g(x_1)-g(x_2)|<\delta \tag{B-15}$$

对每个 $x_1,x_2 \in [a,b]$ 都成立。在 $[a,b]$ 上的一个或多个点不满足这个定义的函数是不连续的。

这个定义意味着对每个 $x_0 \in [a,b]$，存在一个在式(B-1)第一个连续的定义中单一的 $\{\varepsilon,\delta\}$ 对；则 $\delta=\delta(\xi)$在这个区间上一致成立，它仅是 ε 的函数，而不是任意特定的 $x_0 \in [a,b]$ 的函数。如果一个函数在 $[a,b]$ 上一致连续，则它在任意 $x_0 \in [a,b]$ 处连续。尽管这个定义也可用于开区间或半开区间，但可以证明如果 $g(x)$在闭区间 $[a,b]$ 上的每个点都连续，则它是一致连续的。

例 B-7 指数函数 $g(x)=\exp(-\lambda x)$，$(\lambda>0)$，在$\mathcal{R}^+$上是一致连续的。证明如下：对于 $|x_2-x_1|<\varepsilon$

$$\begin{aligned}|g(x_2)-g(x_1)| &= |\exp(-\lambda x_2)-\exp(-\lambda x_1)| = \exp(-\lambda x_1)|\exp(-\lambda(x_2-x_1))-1| \\ &< \exp(-\lambda x_1)|\exp(\lambda\varepsilon)-1| < |\exp(\lambda\varepsilon)-1| = \delta(\varepsilon)\end{aligned} \tag{B-16}$$

其中在最后一个不等式中，我们用到了在 $x_1=0$ 处 $\exp(-\lambda x_1)$取最大值的事实。对每个 $\delta(\varepsilon)>0$，我们可以指定一个 $\varepsilon>0$，使得不等式对每个 x_1，$x_2 \in \mathcal{R}^+$ 都成立。这在图 B-3 中证明了。注意到 $g(x)=\exp(-\lambda x)$在整个实数轴 R 上并不是一致连续的。

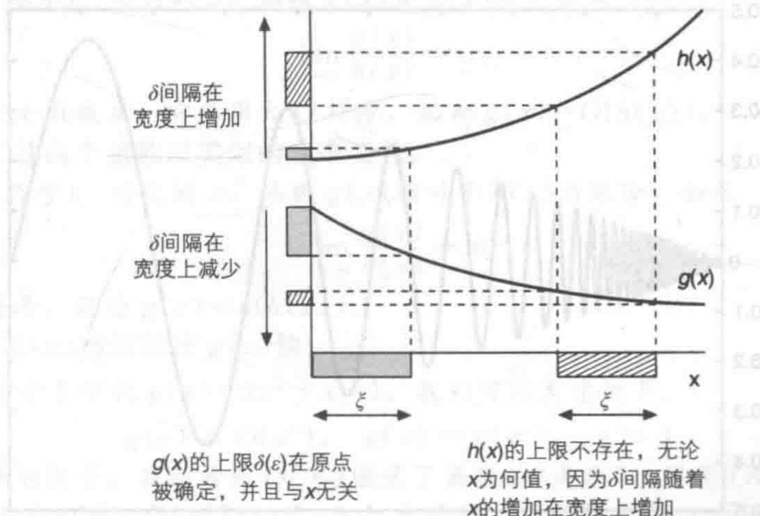


图 B-3　一致连续函数 $g(x)$和不连续函数 $h(x)$的例子，都在$\mathcal{R}^+$上(为了容易看清两个间隔的设置将纵轴在 1 处断开)◀

例 B-8 接下来，考虑 $h(x)=\exp(\lambda x)$，其中 $\lambda>0$。可以直接证明 $h(x)$ 对任意 $x_0\in\mathcal{R}^+$ 是连续的，但不是一致连续的。与式(B-16)类似，对 $|x_2-x_1|<\varepsilon$：

$$\begin{aligned}|h(x_2)-h(x_1)| &= |\exp(\lambda x_2)-\exp(\lambda x_1)| \\ &= \exp(\lambda x_1)|\exp(\lambda(x_2-x_1))-1| \\ &< \exp(\lambda x_1)|\exp(\lambda\varepsilon)-1| = \delta(x_1,\varepsilon)\end{aligned} \tag{B-17}$$

等式右边不能写成与 x_1 无关，图 B-3 证明了 $h(x)$ 在 $\mathcal{R}^+$ 上不是一致连续的。◀

通常，如果存在 $x_1,x_2\in[a,b]$ 使得对小的 ε，$g(x_1)$ 和 $g(x_2)$ 都趋向无穷，则这个函数不是一致连续的(因为 $\infty-\infty$ 是没有定义的)。

例 B-9 函数 $g(x)=1/x$ 在 $[1,\infty)$ 上是一致连续的。观察到由于 x_1，$x_2\geqslant1$，有 $|1/x_1-1/x_2|=|(x_2-x_1)/x_1x_2|\leqslant|x_2-x_1|$。因此，我们可以选 $\delta=\varepsilon$(与 x 无关)，对每个 $\{x_1,x_2\}$ 一致连续条件成立：$|x_1-x_2|<\varepsilon\Rightarrow|1/x_1-1/x_2|<\varepsilon=\delta$。◀

例 B-10 稍后将介绍的单位阶跃函数和符号函数在 $x=0$ 处都不连续，而绝对值函数 $f(x)=|x|$ 在 $[-1,1]$ 上是一致连续的，即使它的一阶导数在 $x=0$ 处符号改变；对于 $|x|$ 中任意 x，不存在"跳变"。符号函数和绝对值函数的相关性如下：

$$\frac{\mathrm{d}}{\mathrm{d}x}|x|=\operatorname{sgn}(x) \tag{B-18}$$

由此式可以清楚地看到 $|x|$ 的导数在 $x=0$ 处不连续。◀

也有另一类连续包括了 $[a,b]$ 上的点序列。

定义(绝对连续)　函数 $g(x)$ 在 $[a,b]$ 上是**绝对连续**的，如果对所有 $\delta>0$ 和有限的 $N\in\mathcal{N}$，存在 $\varepsilon>0$，使得

$$\sum_{n=1}^{N}|x_{1n}-x_{2n}|<\varepsilon\Rightarrow\sum_{n=1}^{N}|g(x_{1n})-g(x_{2n})|<\delta \tag{B-19}$$

其中对于 $n=1,\cdots,N$，$[a,b]$ 的子区间 $\{[x_{1n},x_{2n}]\}$ 是互不相交的。

这类连续需要对任意有限子区间集，δ 只依赖于 ε。注意如果定义了集合 $A\triangleq\bigcup_{n=1}^{N}[x_{1n},x_{2n}]$，则用第 2 章的勒贝格测度可推出式(B-19)的左边等价于 $L(A)<\varepsilon$。

例 B-11 图 B-1b 显示了函数 $g(x)=x^2\sin(2\pi/x)$ 在 $[0,b]$ $(b>0)$ 上是绝对连续的。函数 $g(x)=x\sin(2\pi/x)$ 在 $[0,b]$ 上不是绝对连续的，如图 B-4 所示。

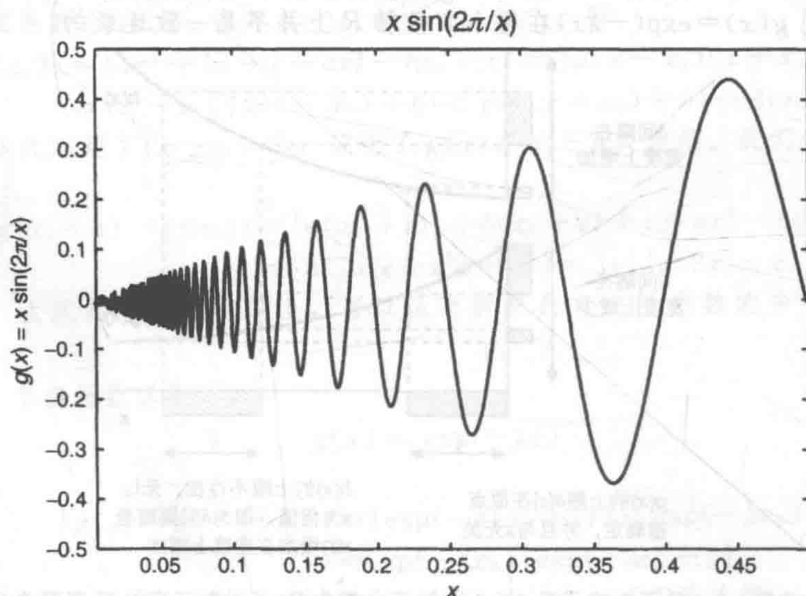


图 B-4　函数 $g(x)=x\sin(2\pi/x)$ 在 $[0,b]$ $(b>0)$ 上不是绝对连续的　◀

定义(有界变分) 函数$g(x)$在$[a,b]$上有**有界变分**，如果存在有限的$M>0$使得

$$\sum_{n=1}^{N}|g(x_n)-g(x_{n-1})|\leqslant M \tag{B-20}$$

对每个排序点集$\{x_0\leqslant x_1\leqslant\cdots\leqslant x_N\}\in[a,b]$和每个有限$N\in\mathcal{N}$都成立。

注意到$a=x_0$和$b=x_N$，除非间隔是开区间或半开区间。可以证明，如果$g(x)$在$[a,b]$上绝对连续，那么它在这个区间上也有有界变分。

例 B-12 在图B-1a和图B-4中已经分别证明了函数$\sin(2\pi/x)$和$x\sin(2\pi/x)$在$[0,b]$上没有有界变分。不管x轴在$x=0$附近怎样变化，对应于式(B-20)中更小的分区，我们会连续看到快速变化；在$[0,b]$上没有式(B-20)中有限的M存在。另一方面，函数$x^n\sin(2\pi/x)$在$[0,b]$有有界变分的条件是$n>1$；图B-1b是$n=2$对应的图。◀

当我们考虑第3章中连续型随机变量，以及讨论第7章随机过程的各类连续时，这个对连续性的概述是非常有用的。本附录的剩下部分，定义了特殊函数并描述了它们的某些属性。

B.2 确界和下确界

定义(确界) 集合$E\in\mathcal{R}$的**确界**是最小上界：它是$\geqslant E$中所有元素的最小的实数。

当分析间隔时确界是非常有用的，因为

$$\sup[a,b)=\sup[a,b]=b \tag{B-21}$$

没有为$E=[a,b)$定义最大值，因为它必须是集合中的一个元素；b也不在这个集合中，对于任何$b-\varepsilon\in E(\varepsilon>0)$，我们总能找到更大的一个$E$的元素。最大值也被称为最大元素。如果一个集合具有最大元素，则确界和最大值是相同的。

定义(下确界) 集合$E\in\mathcal{R}$的**下确界**是最大下界：它是$\leqslant E$中所有元素的最大实数。因此

$$\inf(a,b]=\inf[a,b]=a \tag{B-22}$$

一个集合中的最小值被称为最小元素，没有对如$(a,b]$一样的半开区间定义。如果一个集合具有最小元素，则下确界和最小值是相同的。

B.3 有序符号

有序符号用于表示一个函数$g(x)$当$x\to\infty$时的行为。

定义(大O符号) 对大的x，函数$g(x)$是$h(x)$阶，如果

$$\lim_{x\to\infty}\frac{g(x)}{h(x)}=x_0 \tag{B-23}$$

对于某些非零$x_0\in\mathcal{R}$成立。它采用大O符号，记为$g(x)=O(h(x))$。

这意味着，这两个函数以类似的速率变化。

定义(小O符号) 对大的x，函数$g(x)$相对于$h(x)$为零阶，如果

$$\lim_{x\to\infty}\frac{g(x)}{h(x)}=0 \tag{B-24}$$

采用小o符号，记为$g(x)=o(h(x))$。

这意味着，$h(x)$增加的比$g(x)$快。

例 B-13 对于多项式$g(x)=2x^3+x+1$，我们可以表述如下：

$$g(x)=O(x^3),\quad g(x)=o(x^n),\quad n>3 \tag{B-25}$$

由这个简单的例子，我们看到$O(\cdot)$描述了函数的“类型”。因此$O(1)\Rightarrow a$一个常数，$O(x)\Rightarrow a$一个线性函数，$O(x^2)\Rightarrow a$一个二次函数。其他常见情况是阶乘$O(x!)$、对数$O(\log(x))$和指数$O(a^x)(a>1)$。而小o符号，通常用来表示一个取极限时趋向于零的函数的高阶项(因此可以被忽略)。对以上的函数，如果我们只对当x变大后的三次项感兴

趣，它可以被写做 $g(x)=2x^3+o(x^2)$。

B.4　顶函数和底函数

定义(顶函数)　$x\in\mathcal{R}$的 floor **函数**给出了$\leqslant x$的最大整数：

$$\text{floor}(x)=\lfloor x\rfloor\triangleq\max\{n\in\mathcal{Z}\text{ s.t. }n\leqslant x\}\tag{B-26}$$

定义(底函数)　$x\in\mathcal{R}$的 ceil **函数**给出了$\geqslant x$的最小整数：

$$\text{ceil}(x)=\lceil x\rceil\triangleq\min\{n\in\mathcal{Z}\text{ s.t. }n\geqslant x\}\tag{B-27}$$

B.5　凸函数和凹函数

定义(凸和凹函数)　函数 $g(x)$是**凸**的，如果对任意两个值 x_1，$x_2\in\mathcal{R}$和 $\alpha\in[0,1]$满足：

$$g(\alpha x_1+(1-\alpha)x_2)\leqslant\alpha g(x_1)+(1-\alpha)g(x_2)\tag{B-28}$$

如果 $x_1\neq x_2$，表达式是严格不等式，则函数是严格凸函数。函数 $h(x)$是凹函数，如果$-h(x)$是凸函数。同样，如果对于 $x_1\neq x_2$，$-h(x)$是严格凸函数，则函数 $h(x)$是严格凹函数。

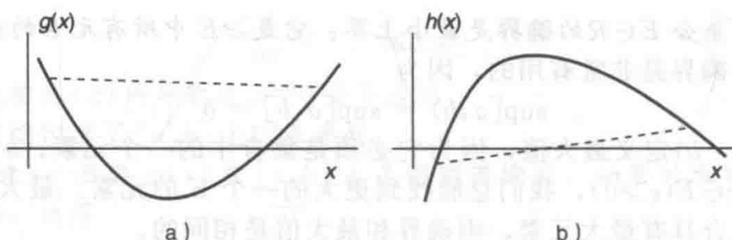


图 B-5　凸函数 $g(x)$和凹函数 $h(x)$的例子。a) 凸函数；b) 凹函数

图 B-5 是凸函数和凹函数的例子。凸函数用于杰森不等式(见附录 F)

B.6　偶函数和奇函数

定义(偶函数和奇函数)　**偶函数** $g_E(x)$和**奇函数** $g_O(x)$由在 $x=0$ 附近具有以下特性来定义：

$$g_E(x)=g_E(-x),\quad g_O(x)=-g_O(-x)\tag{B-29}$$

任意函数 $g(x)$可被分解为：

$$g(x)=g_E(x)+g_O(x)\tag{B-30}$$

其中

$$g_E(x)=[g(x)+g(-x)]/2,\quad g_O(x)=[g(x)-g(-x)]/2\tag{B-31}$$

图 B-6 中给出了这种分解的一个例子。既是偶函数又是奇函数，且满足式(B-31)的唯一函数类型为一个常数函数。

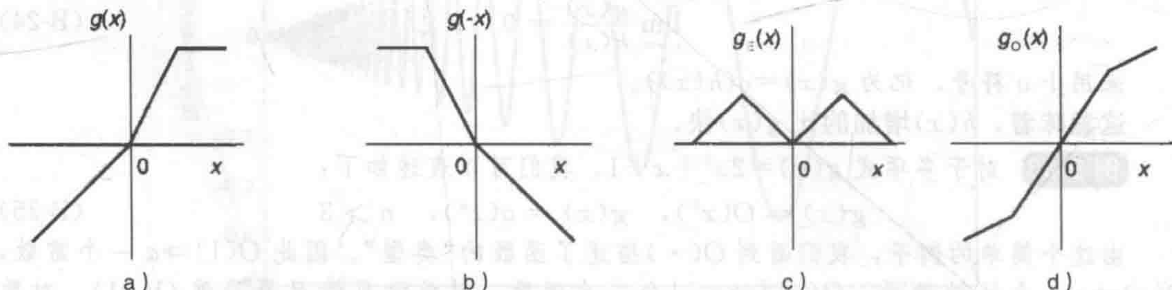


图 B-6　一个分段线性函数的偶和奇分量。a) $g(x)$；b) $g(-x)$；c) $g_E(x)$；d) $g_O(x)$

偶函数和奇函数的其他特性在表 B-1 中不加证明的给出。很容易看出对有限 a:

$$\int_{-a}^{a} g_{\mathrm{O}}(x)\mathrm{d}x = 0 \tag{B-32}$$

表 B-1 偶函数和奇函数的特性

函数	特性	函数	特性
$g_{1\mathrm{E}}(x)g_{2\mathrm{E}}(x)$	偶	$g_{1\mathrm{O}}(x)+g_{2\mathrm{O}}(x)$	奇
$g_{1\mathrm{O}}(x)g_{2\mathrm{O}}(x)$	偶	$g_{\mathrm{E}}(x)+g_{\mathrm{O}}(x)$	两者都不是
$g_{\mathrm{E}}(x)g_{\mathrm{O}}(x)$	奇	$\mathrm{d}g_{\mathrm{E}}(x)/\mathrm{d}x$	奇
$g_{1\mathrm{E}}(x)+g_{2\mathrm{E}}(x)$	偶	$\mathrm{d}g_{\mathrm{O}}(x)/\mathrm{d}x$	偶

这个式子在计算分布的矩时非常有用。当得到以下形式的积分时

$$\int_{-a}^{a} g_{\mathrm{O}}(x)g_{\mathrm{E}}(x)\mathrm{d}x \tag{B-33}$$

我们知道它为零。例如，当计算奇函数的原点矩(见第 5 章)时，当 $f_X(x)$为偶概率密度函数(pdf)(比如零均值高斯随机变量)时，期望值为零。当计算期望值时以下特性也很有用:

$$\int_{-a}^{a} g_{\mathrm{E}}(x)\mathrm{d}x = 2\int_{0}^{a} g_{\mathrm{E}}(x)\mathrm{d}x \tag{B-34}$$

例如，如果 $g_{\mathrm{E}}(x)=|x|$，pdf $f_X(x)$也是偶函数，则我们可以得到:

$$\int_{-a}^{a} |x| f_X(x)\mathrm{d}x = 2\int_{0}^{a} x f_X(x)\mathrm{d}x \tag{B-35}$$

这样我们就可以将绝对值运算去掉了。但如果积分没定义，则对 $a=\infty$ 这些最后的结果不一定成立。柯西随机变量的均值求解就是这种情况。

B.7 符号函数

定义(符号函数) 符号函数为:

$$\operatorname{sgn}(x) \triangleq \begin{cases} -1, & x<0 \\ 0, & x=0 \\ 1, & x>0 \end{cases} \tag{B-36}$$

尽管实际中通常是 $\operatorname{sgn}(0)=1$。

符号函数也可被表示为:

$$\operatorname{sgn}(x) = \frac{\mathrm{d}}{\mathrm{d}x}|x| = x/|x| \tag{B-37}$$

最后一个表达式中 $x\neq 0$，它的微分可写为狄拉克 δ 函数的形式，如下所示:

$$\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{sgn}(x) = 2\delta(x) \tag{B-38}$$

对复数 x 有:

$$\operatorname{sgn}(x) \triangleq (x_{\mathrm{R}}+\mathrm{j}x_{\mathrm{I}})/\sqrt{x_{\mathrm{R}}^2+x_{\mathrm{I}}^2} = x/|x| \tag{B-39}$$

其中 $\{x_{\mathrm{R}},x_{\mathrm{I}}\}$ 是 x 的实部和虚部。注意到 $\operatorname{sgn}(x)\neq\operatorname{sgn}(x_{\mathrm{R}})+\mathrm{j}\operatorname{sgn}(x_{\mathrm{I}})$(即使式子右边是加性高斯白噪声(AWGN)信道中正交相移键控(QPSK)信号的最大似然(ML)检测器：见第10章)。取而代之的是，复变量的符号函数是复平面的单位圆上最接近 x 的点，如图 B-7所示。随着 x 的变化，复数 $\operatorname{sgn}(x)$可以是单位圆上任意一点，而 $\operatorname{sgn}(x_{\mathrm{R}})+\mathrm{j}\operatorname{sgn}(x_{\mathrm{I}})$是仅有的四个点之一：±1±j。

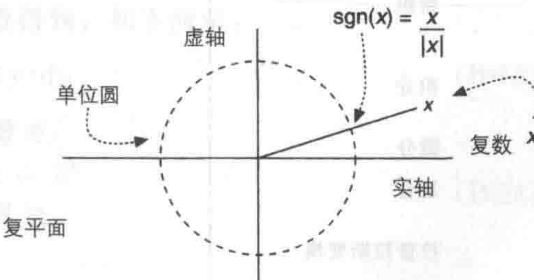


图 B-7 复变量的符号函数：$\operatorname{sgn}(x)$是单位圆上最接近 x 的点

B.8 狄拉克 δ 函数

广义函数是用来描述映射的，理论上说它并不是函数。最知名的广义函数也许是狄拉克 δ 函数。

定义(狄拉克 δ 函数) **狄拉克 δ 函数为**：

$$\delta(x) \triangleq \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0 \end{cases} \quad \text{s.t.} \int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1 \tag{B-40}$$

但是，这个定义不是非常有用。例如，如何将另一函数 $g(x)$ 与 $\delta(x)$ 相乘并不很清晰。实际上我们可以定义如下乘法：

$$g(x)\delta(x-x_0) \triangleq g(x_0)\delta(x-x_0) \tag{B-41}$$

其中，等式右侧仍是在同一位置的 δ 函数，但现在具有面积 $g(x_0)$。这被称为 $\delta(x)$ 的采样特性。在物理系统中，狄拉克 δ 函数可以很方便地描述冲激。我们尤其感兴趣的是，滤波后(通过卷积)和利用积分变换时(如傅里叶变换) $\delta(x)$ 的特性。从采样特性：

$$\int_{-\infty}^{\infty} g(x)\delta(x-x_0)\mathrm{d}x = \int_{-\infty}^{\infty} g(x_0)\delta(x-x_0)\mathrm{d}x$$

$$= g(x_0)\int_{-\infty}^{\infty} \delta(x-x_0)\mathrm{d}x = g(x_0) \tag{B-42}$$

这个结果是 $\delta(x)$ 的筛选特性，因为它“筛选”出了函数 $g(x)$ 在 δ 函数所处位置的值(通过与它相乘)。

我们还可以将狄拉克 δ 函数看作(并定义为)一个单位面积的矩形在宽度趋向于零时的极限结果，其高度趋于无穷：

$$\delta(x) \triangleq \lim_{a\to 0}(1/a)\mathrm{rect}(x/a) \tag{B-43}$$

它同样可以用一个零均值高斯 pdf，当标准差 σ 趋于零来描述：

$$\delta(x) = \lim_{\sigma\to 0} \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-x^2/2\sigma^2) \tag{B-44}$$

对任意 σ，它都为单位面积。后面还会给出它用 sinc 函数表述的定义，如图 B-9 所示。表 B-2 总结了狄拉克 δ 函数的一些有用的属性。卷积和筛选表达式是相同的，因为冲激函数是关于其自变量对称的。狄拉克 δ 函数在定义离散型随机变量的 pdf 时非常有用，因为它可以被积分(例如求矩时)。

表 B-2 狄拉克 δ 函数的属性

属性	表达式
缩放	$\delta(ax) = (1/\lvert a \rvert)\delta(x)(a \neq 0)$
采样	$g(x)\delta(x-x_0) = g(x_0)\delta(x-x_0)$
筛选	$\int_{-\infty}^{\infty} g(x)\delta(x-x_0)\mathrm{d}x = g(x_0)$
卷积	$\int_{-\infty}^{\infty} g(x)\delta(x_0-x)\mathrm{d}x = g(x_0)$
积分	$\int_{-\infty}^{x} \delta(y)\mathrm{d}y = u(x)$
微分	$\delta(x) = \mathrm{d}u(x)/\mathrm{d}x$
复合	$\delta(g(x)) = \delta(x-x_0)/\lvert g'(x_0) \rvert$
拉普拉斯变换	$\int_{-\infty}^{\infty} \delta(x-x_0)\exp(-sx)\mathrm{d}x = \exp(-x_0 s)$

定义(两变量的狄拉克 δ 函数) **二维狄拉克 δ 函数为**：

$$\delta(x,y)=\delta(x)\delta(y)\triangleq\begin{cases}\infty, & x=y=0\\ 0, & \text{其他}\end{cases} \tag{B-45}$$

则

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x,y)\mathrm{d}x\mathrm{d}y=1 \tag{B-46}$$

B.9　克罗内克δ函数

定义(克罗内克δ函数)　**克罗内克δ函数为：**

$$\delta[k]\triangleq\begin{cases}1, & k=0\\ 0, & k\neq 0\end{cases} \tag{B-47}$$

其中$k\in\mathcal{Z}$。

其中用到了方括号，这样克罗内克函数就不会与狄拉克δ函数混淆。(符号δ_k也用于克罗内克δ函数；但我们这里没有使用，因为它类似于第2章中讨论的狄拉克测度。)它具有以下基本特性：

$$\delta[k-n]=\begin{cases}1, & k=n\\ 0, & k\neq n(\text{平移})\end{cases} \tag{B-48}$$

$$\sum_{n=-\infty}^{\infty}\alpha_n\delta[k-n]=\alpha_k(\text{筛选}) \tag{B-49}$$

其中k，$n\in\mathcal{Z}$。克罗内克δ函数在定义概率质量函数(pmf)时非常方便，也可用来表示离散时间信号和系统。

定义(两变量的克罗内克δ函数)　**二维克罗内克δ函数为：**

$$\delta[x,y]=\delta[x]\delta[y]\triangleq\begin{cases}1, & x=y=0\\ 0, & \text{其他}\end{cases} \tag{B-50}$$

B.10　单位阶跃函数

定义(单位阶跃函数)　**单位阶跃函数为：**

$$u(x)\triangleq\begin{cases}1, & x\geqslant 0\\ 0, & x<0\end{cases} \tag{B-51}$$

它也被称为海维塞德阶跃函数，有时用符号$H(x)$表示。

单位阶跃函数也可以用符号函数来定义：

$$u(x)\triangleq[1+\operatorname{sgn}(x)]/2=\begin{cases}0, & x<0\\ 1/2, & x=0\\ 1, & x>0\end{cases} \tag{B-52}$$

由于$x=0$具有勒贝格测度零，通常用式(B-51)中的定义更方便，这是本书主要使用的形式，尤其是当$u(x)$定义随机变量的定义域时。式(B-52)的优势在于它与符号函数定义的一致性，这样它在$x=0$时的值就不再模糊了(尽管前面提到，在实际中经常使用$\operatorname{sgn}(0)=(1)$)。单位阶跃函数也可由狄拉克$\delta$函数得到，如下所示：

$$u(x)=\int_{-\infty}^{x}\delta(y)\mathrm{d}y \tag{B-53}$$

定义(离散单位阶跃函数)　**离散单位阶跃函数为：**

$$u[k]\triangleq\begin{cases}1, & k\in\mathcal{Z}^{+}\\ 0, & \text{其他}\end{cases} \tag{B-54}$$

它可以表示为克罗内克δ函数的形式：

$$u[k]=\sum_{n=0}^{\infty}\delta[k-n] \tag{B-55}$$

B.11 矩形函数

定义(矩形函数) **矩形函数**为：

$$\operatorname{rect}(x) \triangleq \begin{cases} 1, & |x| \leqslant 1/2 \\ 0, & |x| > 1/2 \end{cases} \tag{B-56}$$

有时用符号$\Pi(x)$表示。

它可以用单位阶跃函数表示如下：

$$\operatorname{rect}(x) = u(x+1/2) - u(x-1/2) \tag{B-57}$$

因为$\pm 1/2$的勒贝格测度是零，我们立刻得到$\operatorname{rect}(\pm 1/2) \triangleq 1$，虽然有可能在$x=\pm 1/2$时为零。另一个定义是$x=\pm 1/2$时为$1/2$：

$$\operatorname{rect}(x) \triangleq \begin{cases} 0, & |x| > 1/2 \\ 1/2, & |x| = 1/2 \\ 1, & |x| < 1/2 \end{cases} \tag{B-58}$$

最后这种形式与符号函数的定义一致，所以$\operatorname{sgn}(x)$可通过适当的缩放和移位用$\operatorname{rect}(x)$来表示。基于上述第一个定义的矩形函数通常被用来指定函数的定义域(参见下面的三角形函数)。

定义(离散矩形函数) **离散矩形函数**是：

$$\operatorname{rect}[x] \triangleq \begin{cases} 1, & |x| = 0,1,\ldots,N-1 \\ 0, & \text{其他} \end{cases} \tag{B-59}$$

其中$N \in \mathbb{Z}$。

注意，在数字滤波器设计中，这个函数被称为一个矩形窗。

B.12 三角函数和斜坡函数

定义(三角函数) **三角函数**为：

$$\operatorname{tri}(x) \triangleq (1-|x|)\operatorname{rect}(x/2) = \begin{cases} 1-|x|, & |x| \leqslant 1 \\ 0, & |x| > 1 \end{cases} \tag{B-60}$$

有时也用符号$\Lambda(x)$。

上式通过适当截断$1-|x|$而用矩形函数说明$\operatorname{tri}(x)$的定义域。三角函数可通过两个矩形函数卷积得到：

$$\operatorname{tri}(x) = \int_{-\infty}^{\infty} \operatorname{rect}(x-v)\operatorname{rect}(v)\,\mathrm{d}v \tag{B-61}$$

定义(斜坡函数) **斜坡函数**为：

$$R(x) \triangleq xu(x) = \begin{cases} x, & x \geqslant 0 \\ 0, & x < 0 \end{cases} \tag{B-62}$$

式(B-62)中的单位阶跃函数通过截断线性函数$g(x)=x$定义了$R(x)$的定义域。通过对单位阶跃函数积分得到斜坡函数：

$$R(x) = \int_{-\infty}^{x} u(v)\,\mathrm{d}v \tag{B-63}$$

B.13 指示函数

定义(指示函数) 集合E的**指示函数**为：

$$I_{\mathrm{E}}(x) \triangleq \begin{cases} 1, & x \in E \\ 0, & \text{其他} \end{cases} \tag{B-64}$$

当$E=[a, b]$，$(a<b)$时，指示函数指定了一个闭区间：

$$I_{[a,b]}(x) = \begin{cases} 1, & a \leqslant x \leqslant b \\ 0, & \text{其他} \end{cases} \tag{B-65}$$

开和半开区间的指示函数类似，用 $I_{[a,b)}(x)$，$I_{(a,b]}(x)$和 $I_{(a,b)}(x)$表示。有可能 $a=-\infty$ 或 $b=\infty$或两者都有，这种情况下，用 $I_{\mathcal{R}}(x)$表示。指示函数可以用单位阶跃函数表示如下：

$$I_{[0,\infty)}(x)=u(x), I_{[a,\infty)}(x)=u(x-a),\quad I_{[a,b)}(x)=u(x-a)-u(x-b) \tag{B-66}$$

它也可以被用于可数集 $E\in\mathcal{Z}$如

$$I_{\{0,1,\cdots\}}[x]=I_{\mathcal{Z}^+}[x]=\begin{cases}1, & x=0,1,\cdots\\0, & \text{其他}\end{cases} \tag{B-67}$$

其中方括号用来强调 x 是离散的。指示函数在指定 pdf 的定义域时是非常方便的(对指数型随机变量)：

$$f_X(x)=\lambda\exp(-\lambda x)I_{[0,\infty)}(x) \tag{B-68}$$

或 pmf(对泊松随机变量)：

$$p_X[x]=\exp(-\alpha)(\alpha^x/x!)I_{\mathcal{Z}^+}[x] \tag{B-69}$$

补集的指示函数为：

$$I_{E^c}(x)=\begin{cases}0, & x\in E\\1, & \text{其他}\end{cases}=1-I_E(x) \tag{B-70}$$

对交集为：

$$I_{E\cap F}(x)=\begin{cases}1, & x\in E\cap F\\0, & \text{其他}\end{cases}=I_E(x)I_F(x) \tag{B-71}$$

对于不相交的集合 E 和 F 的并集，我们可以写为 $I_{E\cup F}(x)=I_E(x)+I_F(x)$。然而，如果这两个集合是相交的，必须减去一项，如下所示：

$$I_{E\cup F}(x)=I_E(x)+I_F(x)-I_{E\cap F}(x)=I_E(x)+I_F(x)-I_E(x)I_F(x) \tag{B-72}$$

注意这个表达式与式(2-49)中用于重叠事件概率的表达式的相似性，即 $P(E\cup F)=p(E)+P(F)-P(EF)$。实际上，第5章中描述的指示函数的期望属性：

$$\varepsilon[I_{E\cup F}(x)]=P(x\in E\cup F)=P(x\in E)+P(x\in F)-P(x\in EF) \tag{B-73}$$

B.14 辛格函数

定义(辛格函数) **辛格函数为**：

$$\operatorname{sinc}(x)\triangleq\frac{\sin(\pi x)}{\pi x} \tag{B-74}$$

需要注意的是在符号 sinc(x)中并未将 π 显式地示出。图 B-8 为辛格函数，说明对 $x\in\mathcal{Z}$(除了零值)有其他过零点产生。这种函数还有一种不包括 π 的形式，主要用在数学中，如下所示：

$$\operatorname{Sa}(x)\triangleq\frac{\sin(x)}{x} \tag{B-75}$$

它也被绘制在图 B-8 中。也被称为非标准化的辛格函数，它的过零点发生在除零之外的 π 的整数倍处。

辛格函数与伽马函数(在 B-15 节描述)的关系如下：

$$\operatorname{sinc}(x)=[\Gamma(1+x)\Gamma(1-x)]^{-1} \tag{B-76}$$

辛格函数和矩形函数是傅里叶变换对的关系：

$$\operatorname{sinc}(x)=\int_{-\infty}^{\infty}\operatorname{rect}(f)\exp(\mathrm{j}2\pi fx)\mathrm{d}f \tag{B-77}$$

$$\operatorname{rect}(f)=\int_{-\infty}^{\infty}\operatorname{sinc}(x)\exp(-\mathrm{j}2\pi fx)\mathrm{d}x \tag{B-78}$$

这表明它具有单位面积：

$$\int_{-\infty}^{\infty}\operatorname{sinc}(x)\mathrm{d}x=\operatorname{rect}(0)=1 \tag{B-79}$$

辛格函数也可以用来定义狄拉克 δ 函数，如下所示：

$$\lim_{B\to\infty} B\mathrm{sinc}(Bx) = \delta(x) \tag{B-80}$$

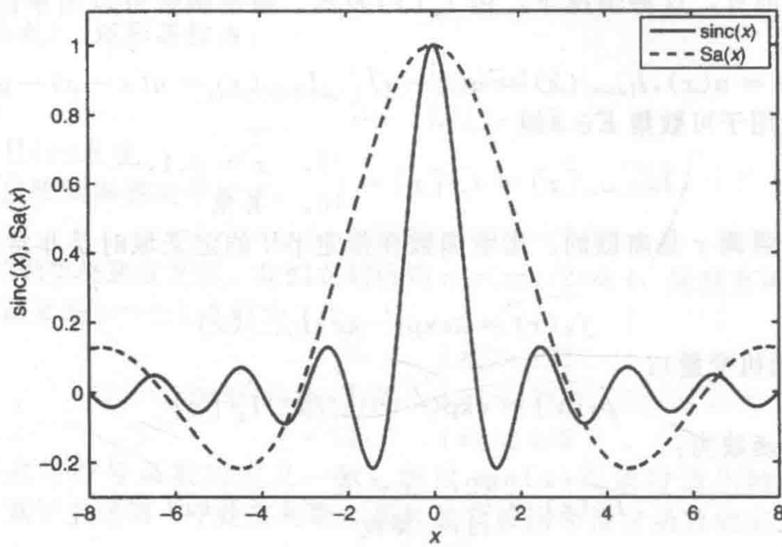


图 B-8 sinc 和 Sa 函数

其中，比例值 B 确保了函数具有单位面积。实例示于图 B-9。

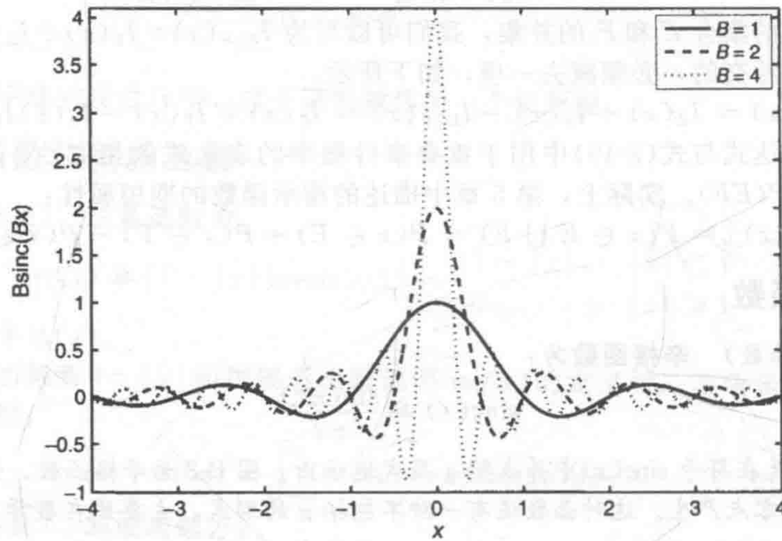


图 B-9 Bsinc(B_x)对应各种 B 值

B.15 对数函数

对数函数用于熵和互信息(见第 10 章)的定义。它也出现在对数随机变量中，可用于随机变量的某些变换。图 B-10 中绘制了自然对数，其中我们看到，它是单调递增函数。习题 10-14 中证明了自然对数是有界的，如下所示：

$$\ln(x) \leqslant x-1 \tag{B-81}$$

图中也证明了这点。这个边界在信息论中证明各种属性时非常有用；习题 10-19 给出了一个例子证明了互信息是非负的。只有当 $x=1$，使得 $\ln(1)=0$ 才能达到上限。注意，这个边界通常并不适用于其他对数底，例如，图中示出了 $\log_2(x)$在某个区间内超过 $x-1$，略高于 $x=1$。我们可以使用下面的公式在不同的对数底间转换：

$$\log_b(x) = \log_b(a)\log_a(x) = \log_a(x)/\log_a(b) \tag{B-82}$$

因此，如果要计算 $\log_2(x)$的上限，可以使用下式：

$$\log_2(x) = \log_2(\mathrm{e})\ln(x) \leqslant 1.4427(x-1) \tag{B-83}$$

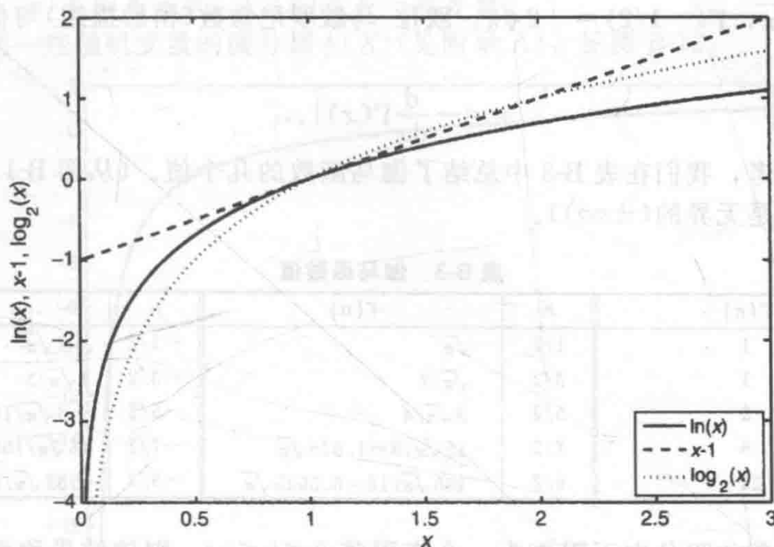


图 B-10　ln(x)的上界

因为无论是斜率还是 $x-1$ 的纵坐标都被比例因子 1.4427 改变了，在某间隔内超过了式(B-83)中的上限，略低于 $x=1$(此上限没有在图中显示)。

B.16　伽马函数

定义(伽马函数)　**伽马函数**为如下定积分：

$$\Gamma(a) \triangleq \int_0^\infty x^{a-1}\exp(-x)\mathrm{d}x \tag{B-84}$$

其中 $a\in\mathcal{C}$(虽然本书中我们只考虑实数)。

它具有以下特性 $\Gamma(a+1)=a\Gamma(a)$，其中 $a=n\in\mathcal{N}$：

$$\Gamma(n)=(n-1)\Gamma(n-1)=(n-1)(n-2)\cdots 2\Gamma(1)=(n-1)! \tag{B-85}$$

对非整数 a，伽马函数可以被视为阶乘函数的扩展。图 B-11 给出了 Γ(a)的图。观察到，对于负整数，函数在±∞之间变化，而对于正整数，它是阶乘函数(幅值下移 1)，由式(B-85)给出。还可以证明，对于 $n\in\mathcal{N}$：

$$\Gamma(1/2+n)=[(2n)!/4^n n!]\sqrt{\pi},$$
$$\Gamma(1/2-n)=[(-4)^n n!/(2n)!]\sqrt{\pi} \tag{B-86}$$

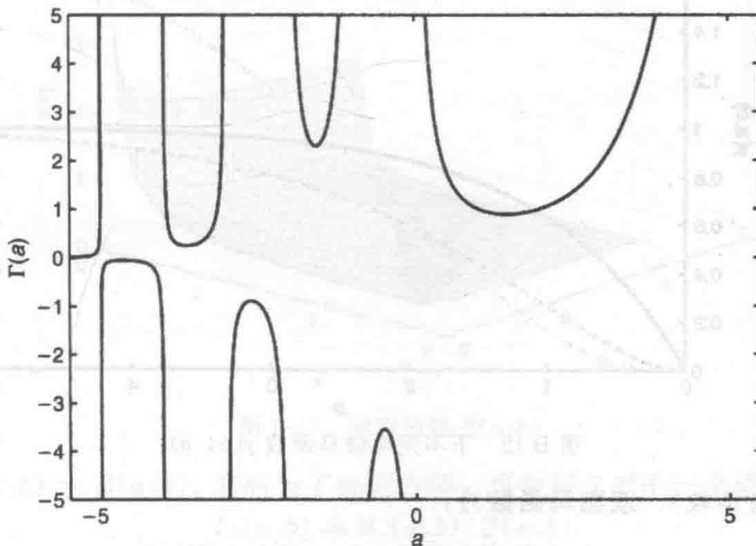


图 B-11　伽马函数 Γ(a)

其中 $\Gamma(1/2)=\sqrt{\pi}$，$\Gamma(-1/2)=-2\sqrt{\pi}$。欧拉-马歇罗尼常数(稍后描述)与伽马函数关系如下所示：

$$\gamma = -\frac{\mathrm{d}}{\mathrm{d}x}\Gamma(x)\big|_{x=1} \tag{B-87}$$

为了便于参考，我们在表 B-3 中总结了伽马函数的几个值。(从图 B-11 回想一下，伽马函数对负整数是无界的($\pm\infty$))。

表 B-3 伽马函数值

n	$\Gamma(n)$	n	$\Gamma(n)$	n	$\Gamma(n)$
1	1	1/2	$\sqrt{\pi}$	−1/2	$-2\sqrt{\pi}$
2	1	3/2	$\sqrt{\pi}/2$	−3/2	$4\sqrt{\pi}/3$
3	2	5/2	$3\sqrt{\pi}/4$	−5/2	$-8\sqrt{\pi}/15\approx-0.5333\sqrt{\pi}$
4	6	7/2	$15\sqrt{\pi}/8=1.875\sqrt{\pi}$	−7/2	$16\sqrt{\pi}/105\approx0.1523\sqrt{\pi}$
5	24	9/2	$105\sqrt{\pi}/16=6.5625\sqrt{\pi}$	−9/2	$-32\sqrt{\pi}/945\approx-0.0339\sqrt{\pi}$

如果伽马函数的积分中下限变为一个有限值 $0<b<\infty$，则该结果称为上不完全伽马函数：

$$\Gamma(a,b) \triangleq \int_b^\infty x^{a-1}\exp(-x)\mathrm{d}x \tag{B-88}$$

其中 a，$b>0$。显然，$\Gamma(a,0)\equiv\Gamma(a)$。下不完全伽马函数为：

$$\gamma(a,b) \triangleq \int_0^b x^{a-1}\exp(-x)\mathrm{d}x \tag{B-89}$$

其中 $\gamma(a,\infty)\equiv\Gamma(a)$。不完全伽马函数由如下递归描述：

$$\Gamma(a,b) = (a-1)\Gamma(a-1,b) + b^{a-1}\exp(-b) \tag{B-90}$$

$$\gamma(a,b) = (a-1)\gamma(a-1,b) - b^{a-1}\exp(-b) \tag{B-91}$$

还需要注意的是 $\Gamma(a)=\Gamma(a,b)+\gamma(a,b)$。下不完全伽马函数在表示一些随机变量的 cdf 很有用，如卡方分布。实例绘于图 B-12。它还具有以下级数展开形式：

$$\gamma(a,b) = b^a\exp(-b)(a-1)!\sum_{n=0}^{\infty}\frac{b^n}{(a+n)!} \tag{B-92}$$

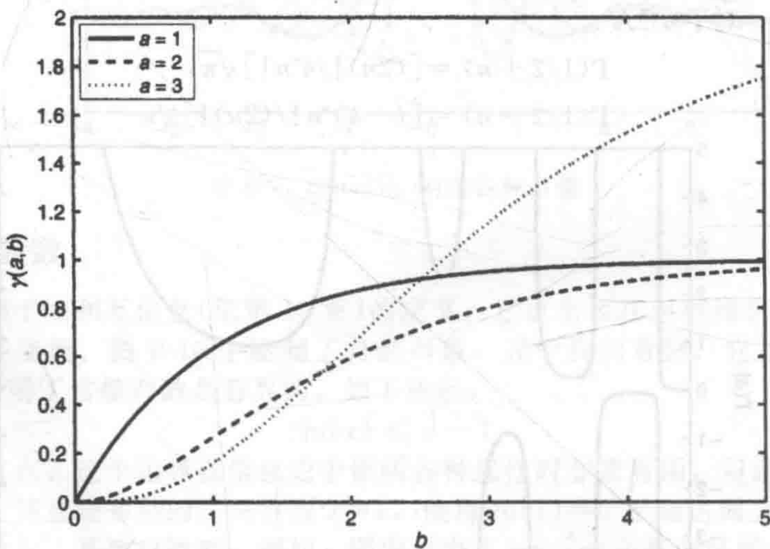


图 B-12 下不完全伽马函数 $\gamma(a,b)$

定义(双伽马函数) **双伽马函数为：**

$$\psi(x) \triangleq \frac{\mathrm{d}}{\mathrm{d}x}\ln(\Gamma(x)) = \Gamma'(x)/\Gamma(x) \tag{B-93}$$

其中 $x \in \mathcal{R}^+$，$\Gamma(x)$是伽马函数，$\Gamma'(x)$是其一般导数。

它用于描述一些随机变量的微分熵 $h(X)$(见附录 A)，见图 B-13。

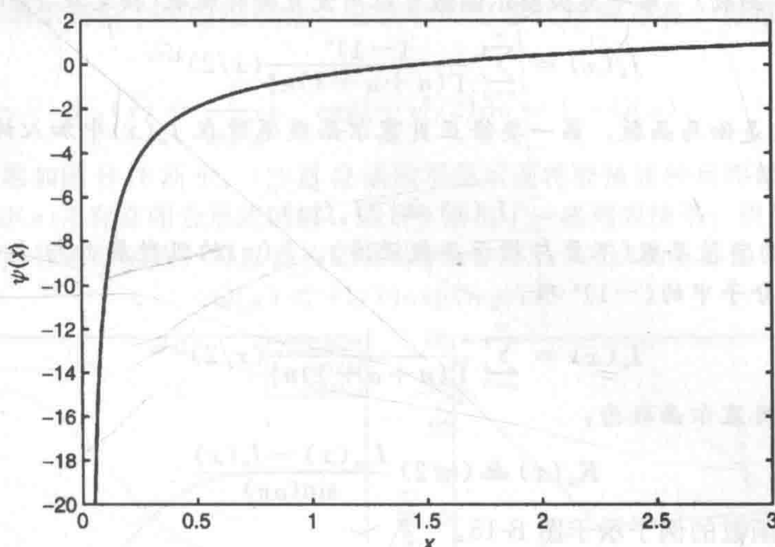


图 B-13　双伽马函数 $\psi(x)$

B.17　贝塔函数

定义(贝塔函数)　**贝塔函数为：**

$$B(a,b) \triangleq \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{B-94}$$

其中 a，$b>0$。

贝塔函数关于其参数对称：$B(a,b)=B(b,a)$，如图 B-14 所示。不完全贝塔函数用不定积分代替定积分得到，如下所示：

$$B_y(a,b) \triangleq \int_0^y x^{a-1}(1-x)^{b-1}\,\mathrm{d}x \tag{B-95}$$

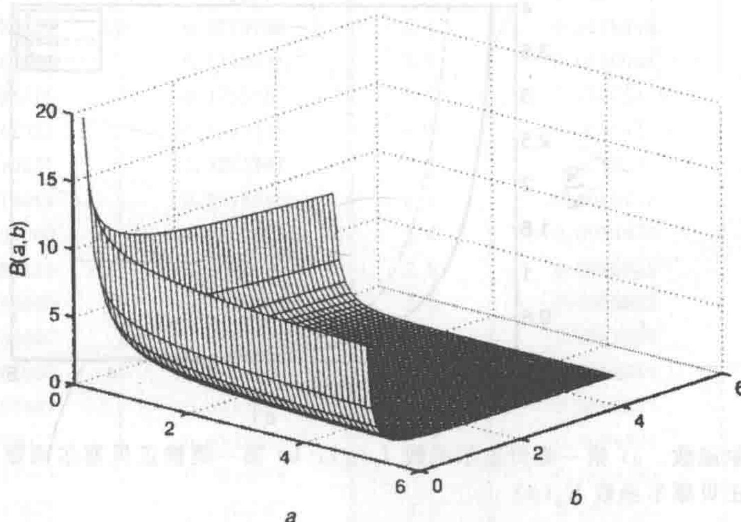


图 B-14　贝塔函数 $B(a,b)$

显然，$B_1(a,b)=B(a,b)$。有时为了标记方便，可使用正则不完全贝塔函数：

$$I_y(a,b) \triangleq B_y(a,b)/B(a,b) \tag{B-96}$$

这是归一化的形式(这里的符号不要与指示函数混淆)。

B.18　贝塞尔函数

定义(贝塞尔函数)　第一类**贝塞尔函数**可以用麦克劳林级数(指定在 $x=0$ 附近)定义为:

$$J_\alpha(x)=\sum_{n=0}^{\infty}\frac{(-1)^n}{\Gamma(n+\alpha+1)n!}(x/2)^{2n+\alpha} \tag{B-97}$$

其中 $\alpha\in\mathcal{R}$，$\Gamma(n)$是伽马函数。第一类修正贝塞尔函数通过在 $J_\alpha(x)$ 中加入纯虚数 $\mathrm{j}=\sqrt{-1}$ 产生，如下所示:

$$I_\alpha(x)\triangleq \mathrm{j}^{-\alpha}J_\alpha(\mathrm{j}x) \tag{B-98}$$

其中 $I_\alpha(x)$是 x 的实值函数(不要与指示函数混淆)。$I_\alpha(x)$的级数展开与 $J_\alpha(x)$的级数展开是相同的，除了分子中的$(-1)^n$ 项:

$$I_\alpha(x)=\sum_{n=0}^{\infty}\frac{1}{\Gamma(n+\alpha+1)n!}(x/2)^{2n+\alpha} \tag{B-99}$$

第二类修正贝塞尔函数为:

$$K_\alpha(x)\triangleq(\pi/2)\frac{I_{-\alpha}(x)-I_\alpha(x)}{\sin(\alpha\pi)} \tag{B-100}$$

各种贝塞尔函数的例子示于图 B-15。

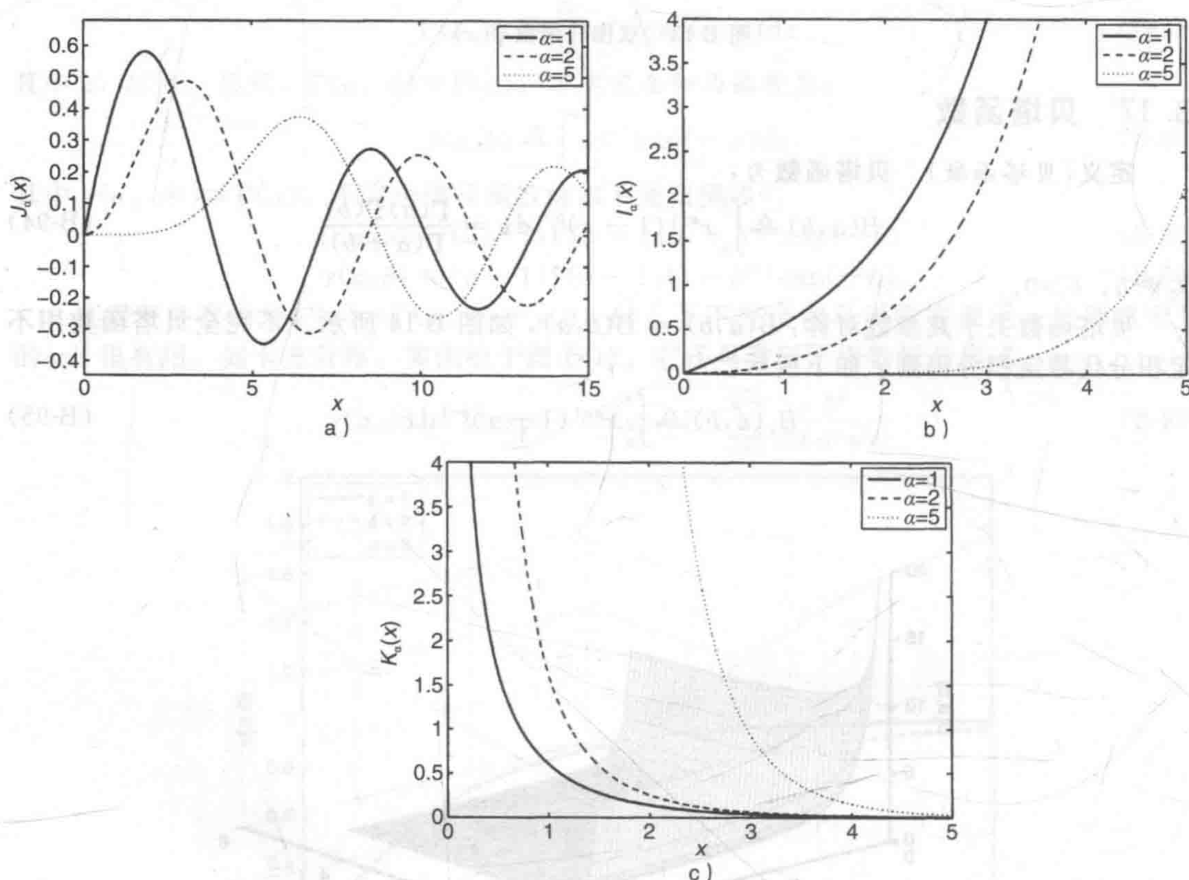


图 B-15　贝塞尔函数。a) 第一类贝塞尔函数 $J_\alpha(x)$；b) 第一类修正贝塞尔函数 $I_\alpha(x)$；c) 第二类修正贝塞尔函数 $K_\alpha(x)$

B.19　Q 函数和误差函数

定义(Q 函数)　Q **函数**定义为标准高斯 pdf 下的面积:

$$Q(x)\triangleq\frac{1}{\sqrt{2\pi}}\int_x^{\infty}\exp(-v^2/2)\mathrm{d}v \tag{B-101}$$

其中 $x \in \mathcal{R}$，$Q(-\infty)=1$，$Q(\infty)=0$，$Q(0)=0.5$

$$Q(-x)=1-Q(x) \tag{B-102}$$

Q-函数用来表示高斯 pdf 的定积分，它没有闭合形式的解存在。标准高斯随机变量的 cdf 可以写成：

$$F_X(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-v^2/2)\mathrm{d}v=1-Q(x) \tag{B-103}$$

Q-函数的图形如图 B-16 所示。(注意 Q-函数不是后面将要描述的马库姆 Q-函数的一种特殊情况。)由于 $Q(x)$不存在闭合形式的解，表 B-4 给出了一些列表结果。但更方便的是具有一个近似 $Q(x)$的闭合解的函数。可以直接证明对于 $x \in \mathcal{R}^+$，$Q(x)$的上界如下所示：

$$Q(x)\leqslant(1/2)\exp(-x^2) \tag{B-104}$$

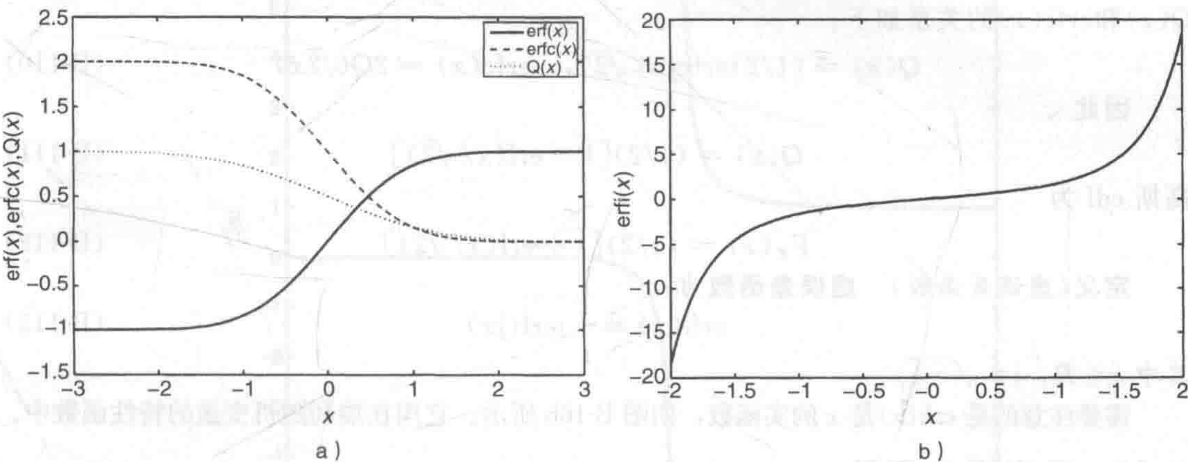


图 B-16　基于高斯 pdf 的函数。a) 误差函数和 Q 函数；b) 虚误差函数

表 B-4　$Q(x)$和 erfc(x)

x	$Q(x)$	erfc(x)	x	$Q(x)$	erfc(x)
0.0	0.5000000	1.0000000	1.8	0.0359303	0.0109095
0.1	0.4601722	0.8875371	1.9	0.0287166	0.0072096
0.2	0.4207440	0.7772974	2.0	0.0227501	0.0046777
0.3	0.3820886	0.6713732	2.1	0.0178644	0.0029795
0.4	0.3445783	0.5716076	2.2	0.0139034	0.0018628
0.5	0.3085375	0.4795001	2.3	0.0107241	0.0011432
0.6	0.2742531	0.3961439	2.4	0.0081975	0.0006885
0.7	0.2419637	0.3221988	2.5	0.0062097	0.0004070
0.8	0.2118554	0.2578990	2.6	0.0046612	0.0002360
0.9	0.0184060	0.2030918	2.7	0.0034670	0.0001343
1.0	0.1586553	0.1572992	2.8	0.0025555	0.0000750
1.1	0.1356660	0.1197949	2.9	0.0018658	0.0000411
1.2	0.1150697	0.0896860	3.0	0.0013499	0.0000221
1.3	0.0968005	0.0659921	3.1	0.0009676	0.0000116
1.4	0.0807567	0.0477149	3.2	0.0006871	0.0000060
1.5	0.0668072	0.0338949	3.3	0.0004834	0.0000031
1.6	0.0547993	0.0236516	3.4	0.0003369	0.0000015
1.7	0.0445655	0.0162095	3.5	0.0002326	0.0000007

定义(误差函数)　**误差函数**是：

$$\mathrm{erf}(x)\triangleq\frac{2}{\sqrt{\pi}}\int_{0}^{x}\exp(-v^2)\mathrm{d}v \tag{B-105}$$

其中 $x \in \mathcal{R}^+$，有 $\mathrm{erf}(0)=0$，$\mathrm{erf}(\infty)=1$ 和

$$\mathrm{erf}(-x)=-\mathrm{erf}(x) \tag{B-106}$$

误差函数是用来表示标准高斯 pdf 下面积的另一类列表函数；它也被绘制在图 B-16a 中。

定义(补余误差函数) 对 $x\in\mathcal{R}$，补余误差函数定义为：

$$\mathrm{erfc}(x)\triangleq 1-\mathrm{erf}(x) \tag{B-107}$$

因此

$$\mathrm{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^{\infty}\exp(-v^2)\mathrm{d}v \tag{B-108}$$

有 $\mathrm{erfc}(\infty)=0$，$\mathrm{erfc}(0)=1$，$\mathrm{erfc}(-\infty)=2$ 和

$$\mathrm{erfc}(-x)=2-\mathrm{erfc}(x) \tag{B-109}$$

它被绘于图 B-16a 中，表 B-4 提供了一些数字值。由式(B-101)和式(B-108)观察到 $Q(x)$和$\mathrm{erfc}(x)$的关系如下：

$$Q(x)=(1/2)\mathrm{erfc}(x/\sqrt{2}),\quad \mathrm{erfc}(x)=2Q(\sqrt{2}x) \tag{B-110}$$

因此

$$Q(x)=(1/2)[1-\mathrm{erf}(x/\sqrt{2})] \tag{B-111}$$

高斯 cdf 为

$$F_X(x)=(1/2)[1+\mathrm{erf}(x/\sqrt{2})] \tag{B-112}$$

定义(虚误差函数) **虚误差函数为**：

$$\mathrm{erfi}(x)\triangleq -\mathrm{j}\,\mathrm{erf}(\mathrm{j}x) \tag{B-113}$$

其中 $x\in\mathcal{R}$，$\mathrm{j}=\sqrt{-1}$。

需要注意的是 $\mathrm{erfi}(x)$是 x 的实函数，如图 B-16b 所示。它用在瑞利随机变量的特性函数中。

B.20 马库姆 Q 函数

定义(马库姆 Q 函数) 广义**马库姆** Q **函数**为：

$$\begin{aligned}Q_m(a,b)&\triangleq\frac{1}{a^{m-1}}\int_b^{\infty}x^m\exp(-(x^2+a^2)/2)I_{m-1}(ax)\mathrm{d}x\\&=\exp(-(a^2+b^2)/2)\sum_{k=1-m}^{\infty}(a/b)^k I_k(ab)\end{aligned} \tag{B-114}$$

其中 $m\in\mathcal{Z}^+$，$\{a,b\}>0$，$I_m(x)$是 m 阶第一类修正贝塞尔函数。

它用在莱斯随机变量的 cdf 中；图 B-17 中绘制的是 $m=1$ 的一个例子。

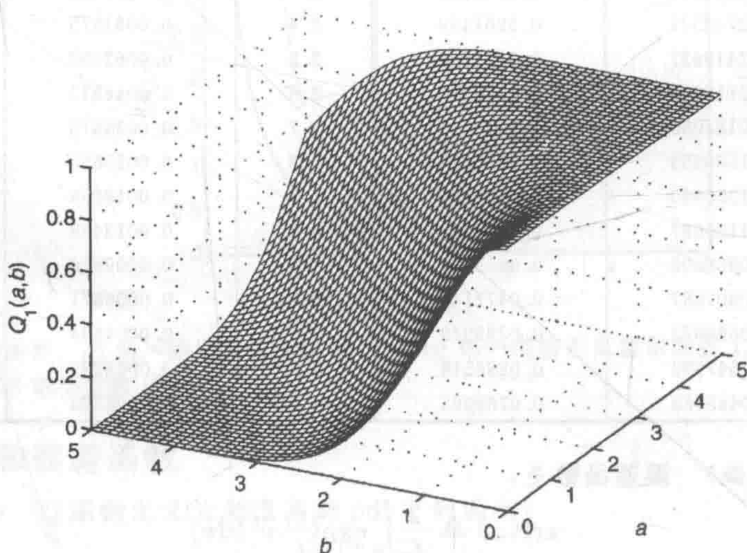


图 B-17 $m=1$ 的马库姆 Q-函数 $Q_m(a,b)$

B. 21 泽塔(ζ)函数

定义(黎曼泽塔函数) **黎曼泽塔函数为**：

$$\zeta(s) \triangleq \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{B-115}$$

其中 $s \triangleq \sigma + j\omega$ 是一个复变量。

只要 $\mathrm{Re}(s)=\sigma>1$，则这个级数收敛。图 B-18 给出了 $\zeta(s)$ 对应实值 s 的图形，对 $\sigma<0$，它收敛到零，对 $\sigma>0$，它收敛到 1，在 $s=1$ 处有一个极点。ζ 函数用在实值 $s\to\alpha$ 的 ζ 随机变量的 pmf 中，它是分布的尺度参数。

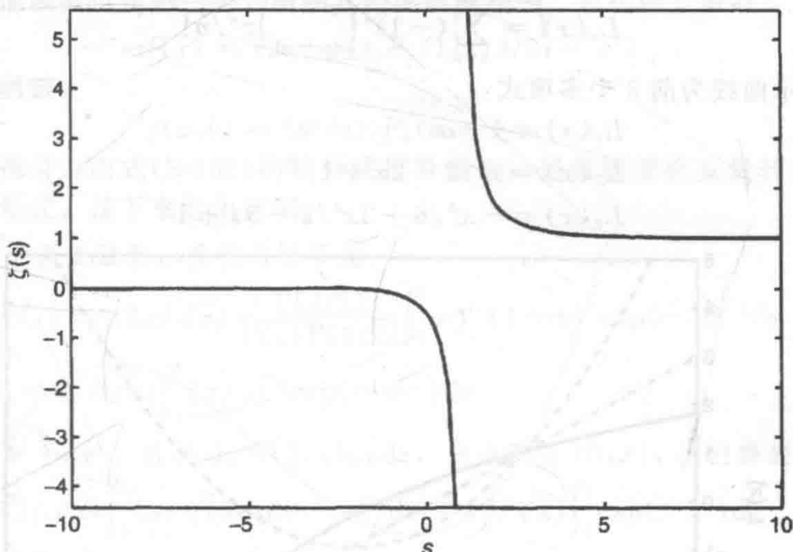


图 B-18 实值 s 的 ζ 函数 $\zeta(S)$

B. 22 递增和递降阶乘

定义(递增和递降阶乘) **递增阶乘为**：

$$x^{(n)} \triangleq x(x+1)\cdots(x+n-1) \tag{B-116}$$

其中 $x\in\mathcal{R}$，$n\in\mathcal{N}$。**递降阶乘为**：

$$(x)_n \triangleq x(x-1)\cdots(x-n+1) \tag{B-117}$$

除了 $x^{(0)}=(x)_0=1$，这些函数产生了 x 的多项式，如

$$(x)_3 = x^3-3x^2+2x, x^{(3)} = x^3+3x^2+2x \tag{B-118}$$

它们的关系如下所示：

$$x^{(n)} = (-1)^n(-x)_n, \quad x^{(n)} = (x+n-1)_n \tag{B-119}$$

具有以下性能：

$$x^{(n)} = \frac{\Gamma(x+n)}{\Gamma(x)}, \quad (x)_n = \frac{\Gamma(x+1)}{\Gamma(x-n+1)}, 1^{(n)} = n!, (n)_n = n! \tag{B-120}$$

$$\frac{x^{(n)}}{n!} = \binom{x+n-1}{n}, \frac{(x)_n}{n!} = \binom{x}{n} \tag{B-121}$$

其中 $\Gamma(x)$ 是伽马函数，$\binom{x}{n}$ 是二项式系数。当 $x\in\mathcal{N}$：

$$x^{(n)} = \frac{(x+n-1)!}{(x-1)!}, \quad (x)_n = \frac{x!}{(x-n)!} \tag{B-122}$$

B.23 拉盖尔多项式

定义(拉盖尔多项式) **拉盖尔多项式**由以下的递归定义：

$$L_{v+1}(x) \triangleq [(2v+1-x)L_v(x) - vL_{v-1}(x)]/(v+1) \tag{B-123}$$

初始条件为 $L_0(x)=1$，$L_1(x)=1-x$。

它们对应于下列微分方程的解，v 是变化的：

$$x\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + (1-x)\frac{\mathrm{d}y}{\mathrm{d}x} + vy = 0 \tag{B-124}$$

拉盖尔多项式的显式表达式为：

$$L_v(x) = \sum_{n=0}^{v}(-1)^n \binom{v}{v-n} x^n/n! \tag{B-125}$$

图 B-19 所示曲线为前 3 个多项式：

$$L_1(x) = 1-x \tag{B-126}$$

$$L_2(x) = x^2/2 - 2x + 1 \tag{B-127}$$

$$L_3(x) = -x^3/6 + 3x^2/2 - 3x + 1 \tag{B-128}$$

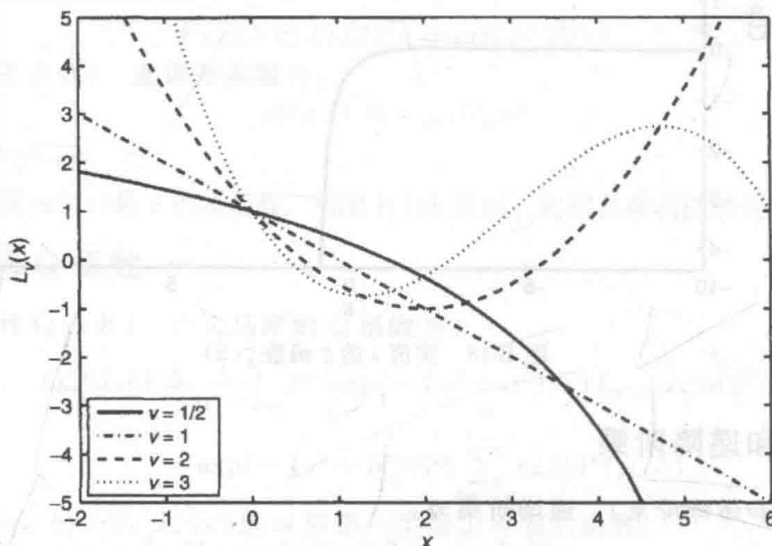


图 B-19 拉盖尔多项式 $L_v(x)$

还包括 $L_{1/2}(x)$，它用在莱斯随机变量的均值和方差中。它可以用第一类修正贝塞尔函数表示如下：

$$L_{1/2}(x) = \exp(x/2)[(1-x)I_0(-x/2) - xI_1(-x/2)] \tag{B-129}$$

拉盖尔多项式，包括小数形式 $L_{n/2}(x)$，可以用 B.24 节所述的超几何函数表示。

B.24 超几何函数

定义(合流超几何函数) 第一类**合流超几何函数**是如下无穷级数：

$$_1F_1(a;b;x) \triangleq \sum_{n=0}^{\infty}\frac{a^{(n)}x^n}{b(n)n!} \tag{B-130}$$

其中$\{a, b\}$是参数，$\{a^{(n)}, b^{(n)}\}$是递增阶乘(见 B.22 节)。对于 a，$b>0$，它由下述积分式表示：

$$_1F_1(a;b;x) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)}\int_0^1 v^{a-1}(1-v)^{b-a-1}\exp(xv)\mathrm{d}v \tag{B-131}$$

其中 $\Gamma(\cdot)$是伽马函数。

这是广义超几何函数 $_pF_q(a;b;x)$（我们没有给出定义，因为本书中没有用到这种通用

形式)的一个特例。${}_1F_1(a;b;x)$ 是下列二阶微分方程的一个解：

$$x\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+(b-x)\frac{\mathrm{d}y}{\mathrm{d}a}-ay=0 \tag{B-132}$$

将该方程与式(B-124)比较，我们看到拉盖尔多项式是一个特例，如下所示：

$$L_v(x)={}_1F_1(-v;1;x) \tag{B-133}$$

用于描述莱斯分布的矩的拉盖尔多项式由 $L_{n/2}(x)={}_1F_1(-n/2;\ 1;\ x)$ 给出。式(B-132)的其他解被称为第二类合流超几何函数，它具有以下符号和积分表示式：

$$\Psi(a;b;x)=\frac{1}{\Gamma(a)}\int_0^\infty v^{a-1}(1+v)^{b-a-1}\exp(-xv)\mathrm{d}v \tag{B-134}$$

还有几个其他函数也是第一类合流超几何函数的特例。包括误差函数：

$$\mathrm{erf}(x)=(2x/\sqrt{\pi})\,{}_1F_1(1/2;3/2;-x^2) \tag{B-135}$$

和下不完全伽马函数：

$$\gamma(a,b)=(b^a/a)\,{}_1F_1(a;a+1;-x) \tag{B-136}$$

这两种式子都可以由式(B-131)的积分式推导出来，通常需要将变量转化以获得所熟悉的特殊函数的形式，接下来给出证明。

例 B-14 对于误差函数，我们可以写为：

$$\begin{aligned}\mathrm{erf}(x)&=(2x/\sqrt{\pi})\frac{\Gamma(3/2)}{\Gamma(1)\Gamma(1/2)}\int_0^1 v^{-1/2}(1-v)^0\exp(-vx^2)\mathrm{d}v\\&=(1/\sqrt{\pi})\int_0^1(x/\sqrt{v})\exp(-vx^2)\mathrm{d}v\end{aligned} \tag{B-137}$$

做变量替换 $z^2=vx^2$，这样 $\mathrm{d}v=(2/x)\sqrt{v}\mathrm{d}z$，积分限为 $\{0,x\}$，我们得到

$$(1/\sqrt{\pi})\int_0^1(x/\sqrt{v})\exp(-vx^2)\mathrm{d}v=(2/\sqrt{\pi})\int_0^x\exp(-z^2)\mathrm{d}z \tag{B-138}$$

正是我们期望的结果。◀

B.25 伯努利数

定义(伯努利数) **伯努利数**$\{B_n\}$是由以下级数的系数指定的有理数：

$$x/[\exp(x)-1]=\sum_{n=0}^{\infty}B_nx^n/n! \tag{B-139}$$

它们包括：$B_0=1$，$B_1=-1/2$，$B_2=1/6$，$B_3=0$，$B_4=-1/30$，$B_5=0$，$B_6=1/42$，$B_7=0$，$B_8=-1/30$，$B_9=0$，$B_{10}=5/66$ 和 $B_{20}=-174611/330$。注意到如果下标可被表示为 $n=2m+1$，其中 $m\in\mathcal{N}$，即对于所有奇数 n(除了 $n=1$)，则 $B_n=0$。伯努利数出现在逻辑随机变量的中心矩的表达式中。

B.26 调和数

定义(调和数) **广义调和数**$\{H_n(v)\}$由以下有限和定义：

$$H_n(v)=\sum_{m=1}^{n}\frac{1}{m^v} \tag{B-140}$$

其中 $v\in(1,\infty)$，$x\in\mathcal{N}$。

当 $n\to\infty$时得到黎曼 ζ 函数。一般的调和数由 $H_n\triangleq H_n(1)$给定，前 10 个值为 $H_1=1$，$H_2=3/2$，$H_3=11/6$，$H_4=25/12$，$H_5=137/60$，$H_6=441/180$，$H_7=3267/1260$，$H_8=6849/2520$，$H_9=64161/22680$，$H_{10}=66429/22680$。调和数与伯努利数关系如下：

$$H_n=\ln(n)+\gamma+(1/2n)-(1/2)\sum_{m=1}^{\infty}B_{2m}n^{-2m}/m \tag{B-141}$$

其中 γ 是 B.27 节描述的欧拉－马歇罗尼常数。从上式，我们看到当 n 很大时$H_n\approx\ln(n)+\gamma$。

B.27 欧拉-马歇罗尼常数

定义(欧拉-马歇罗尼常数) **欧拉-马歇罗尼常数**由以下极限定义：

$$\gamma \triangleq \lim_{n\to\infty}\left[\sum_{m=1}^{n}(1/m)-\ln(n)\right]\approx 0.5772 \tag{B-142}$$

它出现在某些随机变量的微分熵表达式 $h(X)$ 中，以及拉普拉斯变换 $x(t)=\ln(t)\Rightarrow X(s)=-[\ln(s)+\gamma]/s$ 中。

B.28 狄利克函数

定义(狄利克函数) 狄利克函数是：

$$D(x)\triangleq I_{\mathcal{Q}}[x]=\begin{cases}1, & x\in\mathcal{Q}\\ 0, & x\in\mathcal{R}-\mathcal{Q}\end{cases} \tag{B-143}$$

其中 $x\in\mathcal{R}$，$I_{\mathcal{Q}}[x]$是离散的指示函数。

它是一个黎曼不可积函数的例子，是处处不连续的。但是狄利克函数是勒贝格可积的。例如，它在$[0,1]$上的勒贝格积分为零。

进一步阅读

这些函数可在许多数学手册以及各章最后所列参考文献中找到。包括如下：Beyer(1976)，Bracewell(1978)，Gradshteyn 和 Ryzhik(1980)，Kreyszig(1979)，Sokolnikoff 和 Redheffer(1966)，Thomas(1968)，Weisstein(2003)，Zwillinger 和 Kokoska(2000)。

附录 C

频域变换及性质

本附录给出了连续和离散时间函数频域变换的定义，总结了重要性质，还包括几个变换对列表。

C.1 拉普拉斯变换

定义(拉普拉斯变换 $X(s)$) 对 $t\in\mathcal{R}$，$x(t)$的**拉普拉斯变换**为：

$$X(s)=\int_{-\infty}^{\infty}x(t)\exp(-st)\mathrm{d}t \quad (\text{双边}) \tag{C-1}$$

其中 $s\triangleq\sigma+\mathrm{j}\omega\in\mathcal{C}$是单位为 s^{-1} 的复变量。实部 σ 是奈培频率(奈培/秒)，虚部 ω 是角频率(弧度/秒)。收敛域(ROC)指定了 $X(s)$存在条件下 σ 的范围。

式(C-1)是双边拉普拉斯变换。对单边拉普拉斯变换，积分具有有限的下限，通常是零：

$$X(s)\triangleq\int_{0}^{\infty}x(t)\exp(-st)\mathrm{d}t \quad (\text{单边}) \tag{C-2}$$

当 $x(t)$仅在 $t\geqslant 0$ 时非零，单边和双边拉普拉斯变换是相同的，这可以通过写为 $x(t)u(t)$来强调，其中 $u(t)$是单位阶跃函数。拉普拉斯反变换为：

$$x(t)=\int_{\sigma-\mathrm{j}\infty}^{\sigma+\mathrm{j}\infty}X(s)\exp(st)\mathrm{d}s \tag{C-3}$$

其中积分是沿平行于虚轴的线进行的，并且 σ 在 ROC 范围内。但是，当 $X(s)$是一个有理函数时，通常更容易做部分分式展开(PFE)，并由变换对列表来查找 $x(t)$。图 C-1 示出了 s 平面，总结了四类基本函数各自的 ROC。

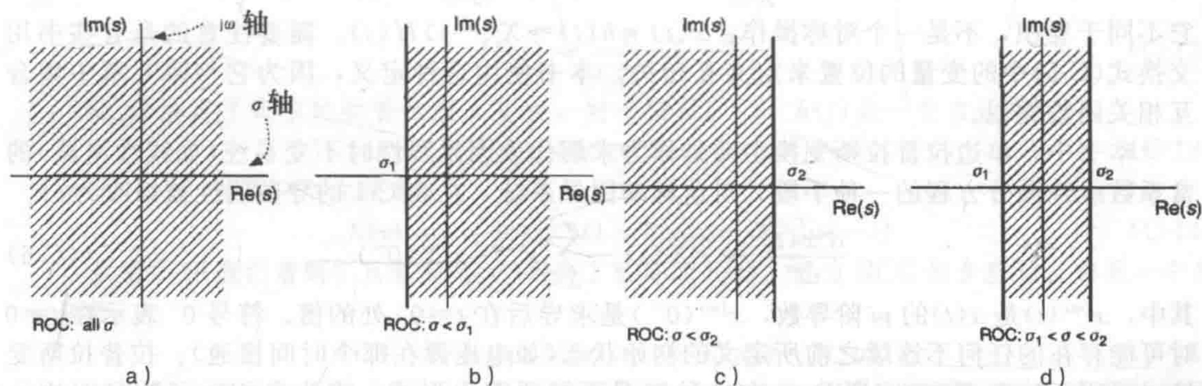


图 C-1 s 平面和 $X(s)$的 ROC。a) 持续时间有限的函数；b) 右函数；c) 左函数；d) 双边函数。如果 ROC 包括虚(jω)轴，函数是有界的(稳定)，这意味着 $\sigma_1<0$，$\sigma_2>0$。如果 $\sigma_1>\sigma_2$，则拉普拉斯变换不存在

表 C-1 总结了单边拉普拉斯变换的几个属性，有的变换对已在表 C-2 中给出。回想一下，两个连续时间函数的卷积为：

$$x(t)*h(t)\triangleq\int_{-\infty}^{\infty}x(\tau)h(t-\tau)\mathrm{d}\tau=\int_{-\infty}^{\infty}h(\tau)x(t-\tau)\mathrm{d}\tau \tag{C-4}$$

表 C-1 拉普拉斯变换的属性

属性	函数	拉普拉斯变换
线性	$a_1x_1(t)+a_2x_2(t)$	$a_1X_1(s)+a_2X_2(s)$
时移性	$x(t-t_0)$	$\exp(-t_0s)X(s)$
尺度特性	$x(\alpha t)$	$(1/\|\alpha\|)X(s/\alpha)$
频移特性	$\exp(s_0t)x(t)$	$X(s-s_0)$
导数	$d^nx(t)/dt^n$	$s^nX(s)$
积分	$\int_{-\infty}^{t}x(\tau)d\tau$	$(1/s)X(s)$
卷积	$x(t)*h(t)$	$X(s)H(s)$
互相关	$x(t)\star h(t)$	$X(-s)H(s)$
乘法	$x(t)h(t)$	$(1/2\pi j)\int_{\sigma-j\infty}^{\sigma+j\infty}X(v)H(s-v)dv$

表 C-2 拉普拉斯变换对

时域 $x(t)$	拉普拉斯变换 $X(s)$	ROC($\sigma=\mathrm{Re}(s)$)
$\delta(t)$	1	$\sigma\in\mathcal{R}$
$t^n\exp(-at)u(t)$	$n!\ /(s+a)^{n+1}$	$\sigma>-a(n\in\mathcal{Z})$
$t^n\exp(-at)u(-t)$	$-n!\ /(s+a)^{n+1}$	$\sigma<-a(n\in\mathcal{Z})$
$\exp(-a\|t\|)$	$2a/(a^2-s^2)$	$-a<\sigma<a$
$\cos(\omega_0t)u(t)$	$s/(s^2+\omega_0^2)$	$\sigma>0$
$\sin(\omega_0t)u(t)$	$\omega_0/(s^2+\omega_0^2)$	$\sigma>0$
$\cosh(at)u(t)$	$s/(s^2-a^2)$	$\sigma>\|a\|$
$\sinh(at)u(t)$	$a/(s^2-a^2)$	$\sigma>\|a\|$
$\exp(-at)\cos(\omega_0t)u(t)$	$(s+a)/[(s+a)^2+\omega_0^2]$	$\sigma>-a$
$\exp(-at)\sin(\omega_0t)u(t)$	$\omega_0/[(s+a)^2+\omega_0^2]$	$\sigma>-a$

这是一个对称操作 $x(t)*h(t)\to X(s)H(s)=H(s)X(s)\to h(t)*x(t)$。它们的互相关为：

$$x(t)\star h(t)\triangleq\int_{-\infty}^{\infty}x(\tau)h(\tau+t)d\tau=\int_{-\infty}^{\infty}x(\tau-t)h(\tau)d\tau \tag{C-5}$$

它不同于卷积，不是一个对称操作：$x(t)*h(t)\to X(-s)H(s)$。需要注意的是有些书用交换式(C-5)中的变量的位置来定义互相关。本书使用这种定义，因为它与第 6 章中集合互相关函数类似。

本书中，单边拉普拉斯变换主要是作为求解代表因果线性时不变系统(如线性电路)的常系数线性微分方程的一种手段。对这样的因果系统，在表 C-1 的导数属性被修改为：

$$\frac{d^nx(t)}{dt^n}\to s^nX(s)-\sum_{m=0}^{n-1}s^{n-m-1}x^{(m)}(0^-) \tag{C-6}$$

其中，$x^{(m)}(t)$是 $x(t)$的 m 阶导数，$x^{(m)}(0^-)$是求导后在 $t=0^-$ 处的值。符号 0^- 表示在 $t=0$ 时可能存在的任何不连续之前所定义的初始状态(如电压源在那个时间接通)。拉普拉斯变换也可用于功率谱密度(PSD)，在这种情况下需要双边形式，因为自相关函数是双边的(它是一个偶函数)。

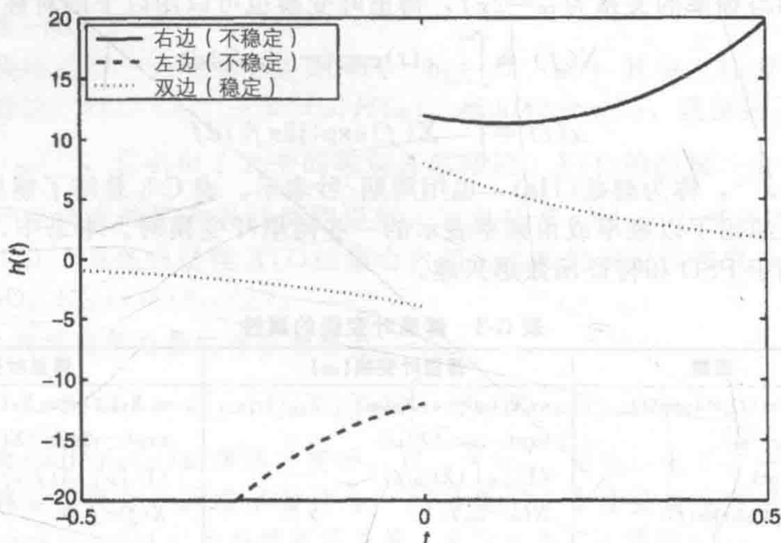
例 C-1 考虑自相关函数：

$$R_{XX}(\tau)=5\exp(-2|\tau|) \tag{C-7}$$

其拉普拉斯变换为以下的 PSD：

$$\begin{aligned}S_{XX}(s)&=5\int_{-\infty}^{0}\exp(2\tau)\exp(-s\tau)d\tau+5\int_{-\infty}^{0}\exp(-2\tau)\exp(-s\tau)d\tau\\&=\frac{5}{2-s}+\frac{5}{s+2}=\frac{20}{4-s^2}\end{aligned} \tag{C-8}$$

它与表C-2中结果匹配。第一个积分对 $\mathrm{Re}(2-s)>0$ 收敛，得到 $\sigma<2$，第二个积分对 $\mathrm{Re}(2+s)>0$ 收敛，得到 $\sigma>-2$，则ROC为 $-2<\sigma<2$。正如预期，它包括虚轴，因为 $R_{XX}(\tau)$ 是一个有界函数。



图C-2 例C-2的冲激响应函数 ◀

例C-2 接下来，我们求以下传递函数对应三个不同ROC的拉普拉斯反变换：(i)$\sigma>3$；(ii)$\sigma<-3$；(iii)$-3<\sigma<3$。

$$H(s)=12\frac{(s-1)}{s^2-9}=\frac{a}{s-3}+\frac{b}{s+3} \tag{C-9}$$

ROC由以下 s 平面上的交集定义：(i)$\{\sigma>-3\}\cap\{\sigma>3\}$；(ii)$\{\sigma<-3\}\cap\{\sigma<3\}$；(iii)$\{\sigma>-3\}\cap\{\sigma<3\}$。对三种情况式(C-9)分子中的系数以同样的方式得到：

$$a=H(s)(s-3)|_{s=3}=24/6=4 \tag{C-10}$$

$$b=H(s)(s+3)|_{s=-3}=48/6=8 \tag{C-11}$$

则

$$H(s)=\frac{4}{s-3}+\frac{8}{s+3} \tag{C-12}$$

ROC决定了每项的拉普拉斯反变换。对于情况(i)，$h(t)$ 是一个右函数：

$$h(t)=[4\exp(3t)+8\exp(-3t)]u(t) \tag{C-13}$$

对情况(ii)，它是一个左函数：

$$h(t)=-[4\exp(3t)+8\exp(-3t)]u(-t) \tag{C-14}$$

从图C-1d我们看到，只有情况(iii)得到了稳定的 $h(t)$，因为ROC包含虚轴。它是一个双边函数：

$$h(t)=-4\exp(3t)u(-t)+8\exp(-3t)u(t) \tag{C-15}$$

将这三个函数都绘制在图C-2中。观察到由于 $h(t)$ 的有界项使得右函数出现轻微下降，但很快无界项成为主导。对于双边函数，$t=0$ 处的值为4，此处有一个不连续。 ◀

C.2 连续时间傅里叶变换

定义(傅里叶变换 $X(\omega)$) $x(t)(t\in\mathcal{R})$的**连续时间傅里叶变换**是：

$$X(\omega)\triangleq\int_{-\infty}^{\infty}x(t)\exp(-\mathrm{j}\omega t)\,\mathrm{d}t \tag{C-16}$$

其中 ω 是角频率。

上式是用 $s=\mathrm{j}\omega$(实部 $\sigma=0$)代入双边拉普拉斯变换 $X(s)$ 得到的，假设ROC包含虚轴

（如图 C-1 所示的四种情况）。傅里叶反变换为：

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)\exp(\mathrm{j}\omega t)\mathrm{d}\omega \tag{C-17}$$

由于角频率与频率的关系为 $\omega=2\pi f$，傅里叶变换也可以用以下的对称变换对表示：

$$X(f)\triangleq\int_{-\infty}^{\infty}x(t)\exp(-\mathrm{j}2\pi ft)\mathrm{d}t \tag{C-18}$$

$$x(t)=\int_{-\infty}^{\infty}X(f)\exp(\mathrm{j}2\pi ft)\mathrm{d}f \tag{C-19}$$

f 的单位是 s^{-1}，称为赫兹(Hz)，也用周期/秒表示。表 C-3 总结了傅里叶变换的几个属性，表 C-4 还给出了以频率或角频率表示的一些傅里叶变换对。本书中，我们主要对傅里叶变换如何用于 PSD 和特征函数感兴趣。

表 C-3 傅里叶变换的属性

属性	函数	傅里叶变换(ω)	傅里叶变换(f)
线性	$a_1x_1(t)+a_2x_2(t)$	$a_1X_1(\omega)+a_2X_2(\omega)$	$a_1X_1(f)+a_2X_2(f)$
时移性	$x(t-t_0)$	$\exp(-\mathrm{j}\omega t_0)X(\omega)$	$\exp(-\mathrm{j}2\pi ft_0)X(f)$
尺度特性	$x(\alpha t)$	$(1/\lvert\alpha\rvert)X(\omega/\alpha)$	$(1/\lvert\alpha\rvert)X(f/\alpha)$
频移特性	$\exp(\mathrm{j}\omega_0 t)x(t)$	$X(\omega-\omega_0)$	$X(f-f_0)$
导数	$\mathrm{d}^nx(t)/\mathrm{d}t^n$	$(\mathrm{j}\omega)^nX(\omega)$	$(\mathrm{j}2\pi f)^nX(f)$
积分	$\int_{-\infty}^{t}x(\tau)\mathrm{d}\tau$	$(1/\mathrm{j}\omega)X(\omega)+\pi X(0)\delta(\omega)$	$(1/\mathrm{j}2\pi f)X(f)+(1/2)X(0)\delta(f)$
卷积	$x(t)*h(t)$	$X(\omega)H(\omega)$	$X(f)H(f)$
互相关	$x(t)\star h(t)$	$X(-\omega)H(\omega)$	$X(-f)H(f)$
自相关	$x(t)\star x(t)$	$\lvert X(\omega)\rvert^2$	$\lvert X(f)\rvert^2$
乘法	$x(t)h(t)$	$(1/2\pi)\int_{-\infty}^{\infty}X(v)H(\omega-v)\mathrm{d}v$	$\int_{-\infty}^{\infty}X(v)H(f-v)\mathrm{d}v$
面积	$x(0)$	$(1/2\pi)\int_{-\infty}^{\infty}X(\omega)\mathrm{d}\omega$	$\int_{-\infty}^{\infty}X(f)\mathrm{d}f$
帕斯瓦尔	$\int_{-\infty}^{\infty}x^2(t)\mathrm{d}t$	$(1/2\pi)\int_{-\infty}^{\infty}\lvert X(\omega)\rvert^2\mathrm{d}\omega$	$\int_{-\infty}^{\infty}\lvert X(f)\rvert^2\mathrm{d}f$

表 C-4 傅里叶变换对

时域 $x(t)$	傅里叶变换 $X(\omega)$	傅里叶变换 $X(f)$
$\delta(t)$	1	1
1	$2\pi\delta(\omega)$	$\delta(f)$
$u(t)$	$\pi\delta(\omega)+1/\mathrm{j}\omega$	$(1/2)\delta(f)+1/\mathrm{j}2\pi f$
$\mathrm{sgn}(t)$	$2/\mathrm{j}\omega$	$1/\mathrm{j}\pi f$
$1/t$	$-\mathrm{j}\pi\mathrm{sgn}(\omega)$	$-\mathrm{j}\pi\mathrm{sgn}(f)$
$\cos(\omega_0 t)$	$\pi[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)]$	$(1/2)[\delta(f-\omega_0/2\pi)+\delta(f+\omega_0/2\pi)]$
$\sin(\omega_0 t)$	$\mathrm{j}\pi[\delta(\omega+\omega_0)-\delta(\omega-\omega_0)]$	$(\mathrm{j}/2)[\delta(f+\omega_0/2\pi)-\delta(f-\omega_0/2\pi)]$
$\exp(-at)u(t)$	$1/(\mathrm{j}\omega+a)(a>0)$	$1/(\mathrm{j}2\pi f+a)(a>0)$
$\exp(-a\lvert t\rvert)$	$2a/(\omega^2+a^2)(a>0)$	$2a/(4\pi^2f^2+a^2)(a>0)$
$\exp(-at^2)$	$\sqrt{\pi/a}\exp(\omega^2/4a)(a>0)$	$\sqrt{\pi/a}\exp(-\pi^2f^2/a)(a>0)$
$\mathrm{rect}(t)$	$\mathrm{sinc}(\omega/2\pi)$	$\mathrm{sinc}(f)$
$\mathrm{sinc}(t)$	$\mathrm{rect}(\omega/2\pi)$	$\mathrm{rect}(f)$
$\mathrm{sinc}^2(t)$	$\mathrm{tri}(\omega/2\pi)$	$\mathrm{tri}(f)$
$\mathrm{tri}(t)$	$\mathrm{sinc}^2(\omega/2\pi)$	$\mathrm{sinc}^2(f)$
$\sum_{m=-\infty}^{\infty}\delta(t-mT)$	$(2\pi/T)\sum_{m=-\infty}^{\infty}\delta(\omega-2\pi m/T)$	$(1/T)\sum_{m=-\infty}^{\infty}\delta(f-m/T)$

例 C-3 例 C-1 中自相关函数的傅里叶变换通过将 $s=\mathrm{j}\omega$ 代入式(C-8)得到：

$$S_{XX}(\omega)=\frac{20}{(\mathrm{j}\omega+2)(2-\mathrm{j}\omega)}=\frac{20}{\omega^2+4} \tag{C-20}$$

这与表 C-4 中结果一致。◀

因为表 C-3 假定函数为实值函数，则 $H(-\omega)=H^*(\omega)$，其中上标表示复共轭。因此，对于互相关特性，$x(t)*h(t)\to X^*(\omega)H(\omega)$。当 $h(t)=x(t)$，就得到了自相关特性：$x(t)\star x(t)\to|X(\omega)|^2$，它引出了表中的帕斯瓦尔结论。$X(t)$的能量，由 $\int_{-\infty}^{\infty}x^2(t)\mathrm{d}t$ 给出，可由其傅里叶变换模的平方下的面积得到，这意味着 $|X(\omega)|^2$ 是能量密度。因此，$x(t)\star x(t)\leftrightarrow|X(\omega)|^2$ 是随机过程 $X(t)$的集合自相关函数的确定性表现形式，它的傅里叶变换是一个 PSD：$R_{XX}(\tau)\leftrightarrow S_{XX}(\omega)$。

例 C-4 一个连续随机变量的特征函数为：

$$\Phi_X(\omega)=\mathcal{E}[\exp(\mathrm{j}\omega X)]=\int_{-\infty}^{\infty}f_X(x)\exp(\mathrm{j}\omega x)\mathrm{d}x \tag{C-21}$$

它是概率密度函数(pdf)$f_X(x)$的傅里叶变换，但 $\omega\to-\omega$。因此，表 C-3 中所有属性都可用于特征函数，将 ω 替换为 $-\omega$(或 f 替换为 $-f$)后表 C-4 中变换对也可用。例如，对于 pdf 为 $f_X(x)=\lambda\exp(-\lambda x)u(x)$的指数随机变量，我们从表 C-4 得到：

$$\Phi_X(\omega)=\lambda/(\lambda-\mathrm{j}\omega) \tag{C-22}$$

但是并没有多少 pdf 是表中的“标准”函数。本书中提到的连续随机变量的特征函数在附录 A 中给出。◀

C.3 z 变换

定义(z 变换 $X(z)$) 对于 $k\in\mathcal{Z}$，$x[k]$的 z 变换为：

$$X(z)\triangleq\sum_{k=-\infty}^{\infty}x[k]z^{-k} \quad (\text{双边变换}) \tag{C-23}$$

其中 $z\triangleq r\exp(\mathrm{j}\omega)$是一个复变量。对 z 平面上的某些 ROC，比如图 C-3 所描绘的四种基本类型序列，上式的和收敛于 $X(z)$。

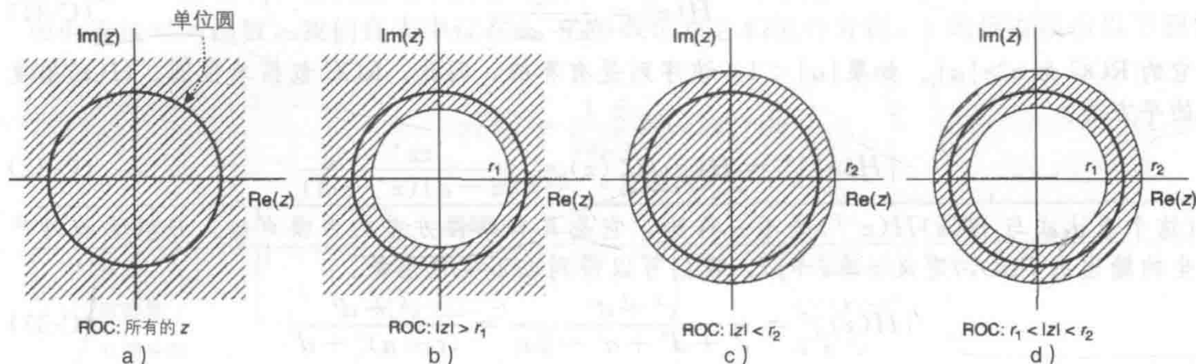


图 C-3 z 平面和 $X(z)$的 ROC。单位圆由 $|z|=1$ 定义。a) 有限持续时间序列；b) 右序列；c) 左序列；d) 双边序列。如果 ROC 包括单位圆，则序列是有界的(稳定的)，这意味着 $r_1<1$ 和 $r_2>1$。如果 $r_1>r_2$ 则 z 变换不存在

第 1 章已介绍了对连续时间信号 $x(t)$均匀采样生成离散时间序列 $x[k]$时所对应的从 s 平面到 z 平面的映射。对单边拉普拉斯变换，积分具有有限的下界，通常为零：

$$X(z)\triangleq\sum_{k=0}^{\infty}x[k]z^{-k} \quad (\text{单边}) \tag{C-24}$$

当 $x[k]$仅在 $k\geqslant0$ 是非零时，单边和双边的 z 变换是相同的，这可以通过写为

$x[k]u[k]$来强调，其中$u[k]$是离散的单位阶跃函数。

例 C-5 考虑下面双边离散时间序列：

$$x[k]=\alpha\beta^{|k|} \tag{C-25}$$

其中α和$|\beta|<1$是实数。双边z变换为

$$X(z)=\alpha\sum_{k=-\infty}^{-1}(\beta z)^{-k}+\alpha\sum_{k=0}^{\infty}(\beta z^{-1})^{k}=\alpha\sum_{m=1}^{\infty}(\beta z)^{m}+\frac{\alpha}{1-\beta z^{-1}} \tag{C-26}$$

其中第一个求和项代入了$m=-k$。第二个求和项用到了附录E中闭合形式的级数的结论，它在$|\beta z^{-1}|<1$即$|z|>|\beta|$时收敛。将第一个求和项加上和减去$m=0$时的α值做如下处理：

$$X(z)=\alpha\sum_{m=0}^{\infty}(\beta z)^{m}-\alpha+\frac{\alpha}{1-\beta z^{-1}}=\frac{\alpha}{1-\beta z}-\alpha+\frac{\alpha}{1-\beta z^{-1}} \tag{C-27}$$

它在$|\beta z|<1$即$|z|<1/|\beta|$时收敛。整体的ROC是$|\beta|<|z|<1/|\beta|$，需要$|\beta|<1$，正如我们的假设。(如果$|\beta|>1$，不存在ROC，因为$|z|<1/|\beta|$和$|z|>|\beta|$不重叠)。ROC包括单位圆$|z|=1$，这对于一个有界序列是必须的。将以上各项合并，我们最终得到

$$X(z)=\frac{\alpha(1-\beta z^{-1})-\alpha(1-\beta z^{-1})(1-\beta z)+\alpha(1-\beta z)}{(1-\beta z^{-1})(1-\beta z)}$$

$$=\frac{\alpha(1-\beta^{2})}{(1-\beta z^{-1})(1-\beta z)} \tag{C-28}$$

这对计算PSD(见第8章)非常有用。特别是对于以下的PSD：

$$S_{XX}(z)=\frac{b}{(1-az^{-1})(1-az)} \tag{C-29}$$

我们可以马上写出对应的自相关函数：

$$R_{XX}[m]=\frac{ba^{|m|}}{1-a^{2}} \tag{C-30}$$

这种类型的PSD在维纳滤波中经常出现(参见第11章)。 ◀

例 C-6 考虑以下右(因果)序列$h[k]=a^{k}u[k]$的z变换，其中a为一个实值：

$$H(z)=\frac{1}{1-az^{-1}} \tag{C-31}$$

它的ROC为$z>|a|$。如果$|a|<1$，该序列是有界的，因此，ROC包括单位圆。$H(z)$幅度的平方为：

$$|H(z)|^{2}=H(z)H^{*}(z)=\frac{zz^{*}}{(z-a)(z^{*}-a)} \tag{C-32}$$

(这个表达式与$H(z)H(z^{-1})$是不一样的，它是具有单位方差的白噪声输入线性系统时产生的输出的PSD。)定义$z\triangleq c+\mathrm{j}d$，我们可以得到以下实值函数：

$$|H(z)|^{2}=\frac{c^{2}+d^{2}}{c^{2}+d^{2}+a^{2}-2ac}=\frac{c^{2}+d^{2}}{(c-a)^{2}+d^{2}} \tag{C-33}$$

通常我们不讨论式(C-33)中的形式，因为它是两个变量(c和d)的函数，如最后一个表达式所示(它也可以被写为半径r和频率ω的函数)。相反，我们对离散时间傅里叶变换(DTFT)的幅度平方$|H(\mathrm{j}\omega)|^{2}$进行检测，因为它是单变量$\omega$的函数(在C.3节定义)。使用式(C-33)中的形式是因为我们要查看图C-4a所示三维图中的极点($z=\alpha=0.9$)和零点($z=0$)。(为了便于查看，我们用了对数坐标。通过选择坐标轴的分辨率使得c接近$\mathrm{Re}(z)=0$和$\mathrm{Re}(z)=0.9$，但不很精确，当然要使c等于这两个值，$10\log|H(z)|^{2}$要分别趋向于$-\infty$和∞，这在图上画不出来)。图C-4b中的函数为$|H(z)|^{2}$，沿实轴$c=\mathrm{Re}(z)$($d=0$)计算，这是对前一个图的"切片"。DTFT $H(\mathrm{j}\omega)$的幅值平方如图C-4c所示。它是通过计算单位圆上的z变换得到的：$z=\exp(\mathrm{j}\omega)$(半径$r=1$)。在$z=0.9$处的极点使得频率响应具

有低通滤波器的特性。

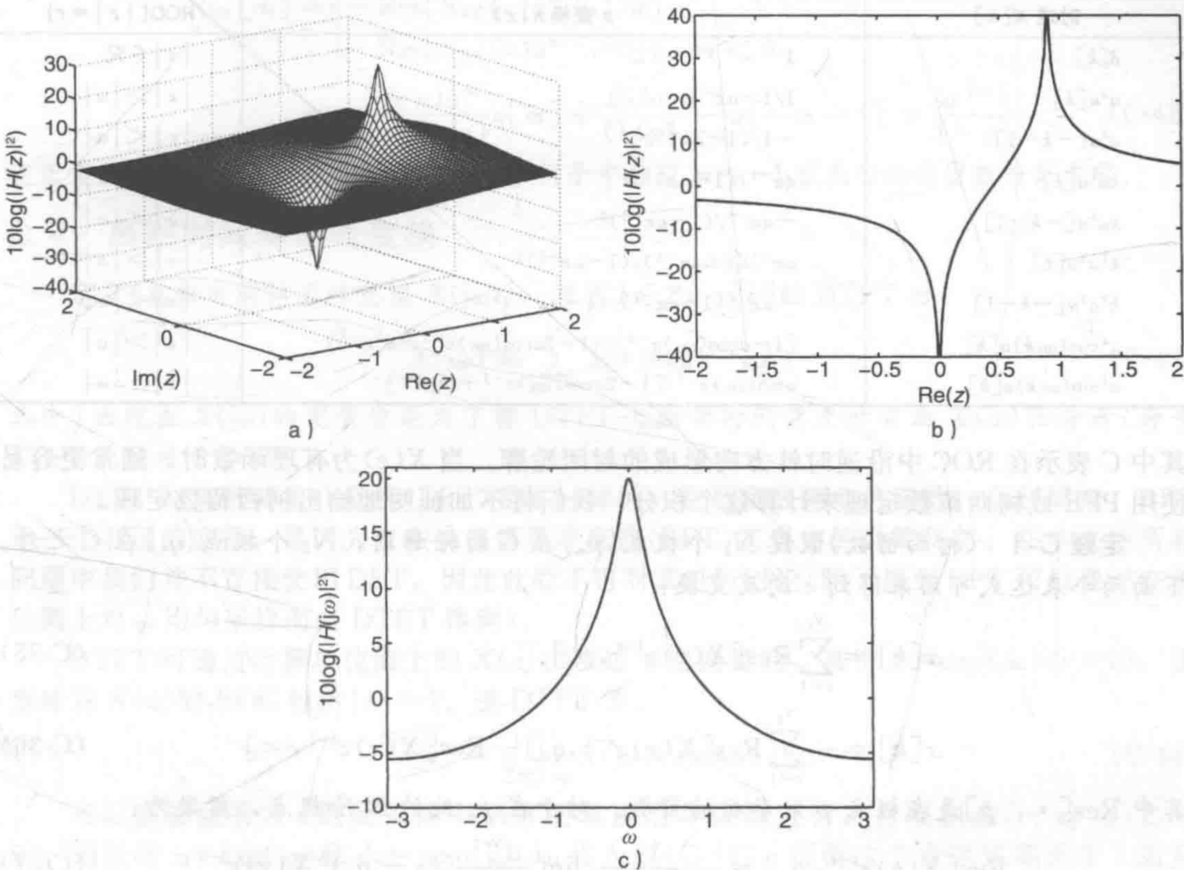


图 C-4 例 C-6 中有一个实数零点($z=0$ 处)和一个实数极点($z=\alpha=0.9$)的一阶系统的 z 变换。a) z 平面上的 $10\log|H(z)|^2$；b) 实轴上的 $10\log|H(z)|^2$（对应 $z=c+\mathrm{j}d$ 中的 c）；c) 在单位圆上的 $10\log|H(z)|^2$（对应 $r=1$ 时 $z=\exp(\mathrm{j}\omega)$ 的 ω）。$H(\mathrm{j}\omega)$ 是 DTFT

表 C-5 总结了 z 变换的几个属性，表 C-6 给出了一些变换对。注意到正弦和余弦是一般的连续时间函数，我们在表中仅在 ω_0 的整数倍对它们进行分析。z 的反变换由以下积分给出：

$$x[k]=\frac{1}{2\pi\mathrm{j}}\oint_C X(z)z^{k-1}\mathrm{d}z \tag{C-34}$$

表 C-5 z 变换的属性

属性	函数	z 变换
线性	$a_1x_1[k]+a_2x_2[k]$	$a_1X_1(z)+a_2X_2(z)$
时移性	$x[k-k_0]$	$z^{-k_0}X(z)$
时间反向	$x[-k]$	$X(z^{-1})$
几何加权	$a^kx[k]$	$X(za^{-1})$
时域加权	$kx[k]$	$-z\mathrm{d}X(z)/\mathrm{d}z$
卷积	$x[k]*h[k]$	$X(z)H(z)$
互相关	$x[k]\star h[k]$	$X(z^{-1})H(z)$
乘法	$x[k]h[k]$	$(1/2\pi\mathrm{j})\oint_C X(v)H(zv^{-1})v^{-1}\mathrm{d}v$
帕斯瓦尔	$\sum_{k=-\infty}^{\infty}x[k]h[k]$	$(1/2\pi j)\oint_C X(v)H(v^{-1})v^{-1}\mathrm{d}v$

表 C-6 z 变换对

时域 $X[k]$	z 变换 $X(z)$	ROC($\|z\|=r$)
$\delta[k]$	1	$\|z\|\in\mathcal{R}$
$a^k u[k]$	$1/1-az^{-1})$	$\|z\|>\|a\|$
$a^k u[-k-1]$	$-1/(1-az^{-1})$	$\|z\|<\|a\|$
$ka^k u[k]$	$az^{-1}/(1-az^{-1})^2$	$\|z\|>\|a\|$
$ka^k u[-k-1]$	$-az^{-1}/(1-az^{-1})^2$	$\|z\|<\|a\|$
$k^2a^k u[k]$	$az^{-1}(1+az^{-1})/(1-az^{-1})^3$	$\|z\|>\|a\|$
$k^2a^k u[-k-1]$	$-az^{-1}(1+az^{-1})/(1-az^{-1})^3$	$\|z\|<\|a\|$
$a^k\cos(\omega_0 k)u[k]$	$[1-a\cos(\omega_0)z^{-1}]/(1-2a\cos(\omega_0)z^{-1}+a^2z^{-2}]$	$\|z\|>\|a\|$
$a^k\sin(\omega_0 k)u[k]$	$a\sin(\omega_0)z^{-1}/[1-2a\cos(\omega_0)z^{-1}+a^2z^{-2}]$	$\|z\|>\|a\|$

其中 C 表示在 ROC 中沿逆时针方向形成的封闭轮廓。当 $X(z)$为有理函数时，通常更容易使用 PFE 或柯西留数定理来计算这个积分，我们将不加证明地给出柯西留数定理。

定理 C-1 （柯西留数）假设 N_1 个极点$\{p_n\}$在 C 的轮廓内，N_2 个极点$\{q_n\}$在 C 之外。下面两个表达式可用来得到 z 的反变换：

$$x[k]=\sum_{n=1}^{N_1}\mathrm{Res}[X(z)z^{k-1},p_n] \tag{C-35}$$

$$x[k]=-\sum_{n=1}^{N_2}\mathrm{Res}[X(z)z^{k-1},q_n]-\mathrm{Res}[X(z)z^{k-1},\infty] \tag{C-36}$$

其中 Res[·，p]是在极点 p 处参数的留数。对于在 p_n 处的 N 阶极点，留数为：

$$\mathrm{Res}[X(z)z^{k-1},p_n]=\frac{1}{(N-1)!}\lim_{z\to p_n}\frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}}\{(z-p_n)^N X(z)z^{k-1}\} \tag{C-37}$$

对 p_n 处一阶极点，上式简化为以下的表达式：

$$\mathrm{Res}[X(z)z^{k-1},p_n]=\lim_{z\to p_n}(z-p_n)X(z)z^{k-1} \tag{C-38}$$

无穷远处的一阶零点的留数为

$$\mathrm{Res}[X(z)z^{k-1},\infty]=-\lim_{z\to\infty}zX(z)z^{k-1} \tag{C-39}$$

无穷远处的高阶零点的留数为零。

注意到 $z=\infty$处的零点相当于 $z=0$ 处的极点，如例 C-7 所示。 ◀

例 C-7 再次考虑式(C-29)中的 PSD，其中 $a<1$ 为实值，ROC 为 $a<|z|<1/a$，如图 C-4d 所示，其中 $r_1=a$，$r_2=1/a$。在 a 处的极点与自相关函数 $R_{XX}[m]$的右边关联，在 $1/a$的极点则与自相关函数的左侧相关联。使用柯西留数定理，我们注意到对 m 的某些值更容易计算在 a 处的留数，而对于其他值，优先计算在 $1/a$ 处的留数。首先，将 PSD(乘以 z^{m-1})重新排列为以下的“标准形式”：

$$S_{XX}(z)z^{m-1}=-\frac{(b/a)z^m}{(z-a)(z-1/a)} \tag{C-40}$$

由分子中 z^m 决定的极点的数目随 m 改变，它反过来又决定了哪个留数更容易计算：

$$m\begin{cases}=0\Rightarrow \text{没有附加的极点} \Rightarrow \text{在 } a \text{ 或 } 1/a \text{ 处计算留数}\\ >0\Rightarrow \text{在 } \infty \text{ 处有附加的极点} \Rightarrow \text{在 } a \text{ 处计算留数}\\ <0\Rightarrow \text{在 } 0 \text{ 处有附加的极点} \Rightarrow \text{在 } 1/a \text{ 处计算留数}\end{cases} \tag{C-41}$$

因为 $m>0$ 不会在单位圆内引入任何附加的极点，式(C-35)可用于计算在 a 处单一极点的留数。如果式(C-35)用于 $m<0$ 时，我们可能需要计算在单位圆内附加的极点(在 $z=0$ 处)的留数。此时应该用式(C-36)，它通常也要求我们计算在无穷远处的一阶零点的留数。因此

$$R_{XX}[m]=\begin{cases}\text{Res}[S_{XX}(z)z^{m-1},a], & m\geqslant 0\\ -\text{Res}[S_{XX}(z)z^{m-1},1/a]\\ -\text{Res}[S_{XX}(z)z^{m-1},\infty], & m<0\end{cases}$$

$$=-\frac{(b/a)a^m}{a-(1/a)}u[m]+\frac{(b/a)a^{-m}}{(1/a)-a}u[-m-1]=\frac{ba^{|m|}}{1-a^2} \qquad \text{(C-42)}$$

这正是式(C-30)中的结果。注意到在这个例子中对应 $m=-1$ 在无穷远的留数恰好为零。◀

C.4 离散时间傅里叶变换

定义(离散时间傅里叶变换 $X(\mathrm{j}\omega)$) 对于 $k\in\mathcal{Z}$，$x[k]$的 DTFT 为：

$$X(\mathrm{j}\omega)\triangleq\sum_{k=-\infty}^{\infty}x[k]\exp(-\mathrm{j}\omega k) \qquad \text{(C-43)}$$

其中 j 出现在 $X(\mathrm{j}\omega)$的变量中是为了将 DTFT 与连续时间傅里叶变换 $X(\omega)$区分开(符号 $X(\mathrm{e}^{\mathrm{j}\omega})$也被用于 DTFT)。

DTFT 与离散傅里叶变换(DFT)是不一样的，它是离散频率的函数。众所周知，DFT 是一个有用的变换，是因为它在快速傅里叶变换(FFT)实现中的计算优势。在本书的所有问题中我们并不直接使用 DFT，因此此处不再对其进行讨论(除了提到 DFT 可以通过在单位圆上对 ω 均匀采样而由 DTFT 得到)。

DTFT 可通过计算单位圆上的 $X(z)$由双边 z 变换得到，其中 $z=\exp(\mathrm{j}\omega)$ $(r=1)$，这意味着 $X(z)$的 ROC 包括$|z|=1$。逆 DTFT 为：

$$x[k]=\frac{1}{2\pi}\int_{-\pi}^{\pi}X(\mathrm{j}\omega)\exp(\mathrm{j}\omega k)\mathrm{d}\omega \qquad \text{(C-44)}$$

可以直接证明当 z 的反变换中积分的轨迹沿单位圆进行，且限制在$[-\pi,\pi]$之间时，通过将 $z=\exp(\mathrm{j}\omega)$和 $\mathrm{d}z=\mathrm{j}\exp(\mathrm{j}\omega)\mathrm{d}\omega$ 代入式(C-34)，可使这个表达式等价于 z 的反变换。表 C-7 总结了 DTFT 的几个属性，表 C-8 给出了一些变换对。注意到 rect$[x]$是附录 B 中定义的离散矩形函数，对于 $k\in\{-(N-1),\cdots,(N-1)\}$ 它是非零的。sinc(Wk) 是通常的连续时间函数，但在表中只计算 W 的整数倍处的值，它的 DTFT 是连续的矩形函数。与连续函数的傅里叶变换一样，我们主要对 DTFT 如何用于 PSD 和特征函数有兴趣。

表 C-7 DTFT 的属性

属性	函数	DTFT
线性	$a_1x_1[k]+a_2x_2[k]$	$a_1X_1(\mathrm{j}\omega)+a_2X_2(\mathrm{j}\omega)$
时移性	$x[k-k_0]$	$\exp(-\mathrm{j}\omega k_0)X(\mathrm{j}\omega)$
时间反向	$x[-k]$	$X(-\mathrm{j}\omega)$
几何加权	$a^kx[k]$	$X(\mathrm{j}\omega/a)$
时域加权	$kx[k]$	$\mathrm{j}\mathrm{d}X(\mathrm{j}\omega)/\mathrm{d}\omega$
卷积	$x[k]*h[k]$	$X(\mathrm{j}\omega)H(\mathrm{j}\omega)$
互相关	$x[k]\star h[k]$	$X(-\mathrm{j}\omega)H(\mathrm{j}\omega)$
自相关	$x[k]\star x[k]$	$\lvert X(\mathrm{j}\omega)\rvert^2$
乘法	$x[k]h[k]$	$(1/2\pi)\int_{-\pi}^{\pi}X(\mathrm{j}v)H(\mathrm{j}(\omega-v))\mathrm{d}v$
面积	$x[0]$	$(1/2\pi)\int_{-\pi}^{\pi}X(\mathrm{j}\omega)\mathrm{d}\omega$
帕斯瓦尔	$\sum_{k=-\infty}^{\infty}x^2[k]$	$(1/2\pi)\int_{-\pi}^{\pi}\lvert X(\mathrm{j}\omega)\rvert^2\mathrm{d}\omega$

表 C-8 DTFT 变换对

时域 $x[k]$	DTFT $X(j\omega)$
$\delta[k]$	1
1	$2\pi\sum_{m=-\infty}^{\infty}\delta(\omega+2\pi m)$
$u[k]$	$1/[1-\exp(-j\omega)]+\pi\sum_{m=-\infty}^{\infty}\delta(\omega+2\pi m)$
$a^k u[k]$	$1/[1-a\exp(-j\omega)](\lvert a\rvert<1)$
$(k+1)a^k u[k]$	$1/[1-a\exp(-j\omega)]^2$
$\cos(\omega_0 k)$	$\pi\sum_{m=-\infty}^{\infty}[\delta(\omega-\omega_0+2\pi m)+\delta(\omega+\omega_0+2\pi m)]$
$\sin(\omega_0 k)$	$j\pi\sum_{m=-\infty}^{\infty}[\delta(\omega+\omega_0+2\pi m)-\delta(\omega-\omega_0+2\pi m)]$
$\mathrm{rect}[k]$	$\sin((N-1/2)\omega)/\sin(\omega/2)$
$\mathrm{sinc}(Wk)$	$(1/W)\mathrm{rect}(\omega/2\pi W)(0<W<1)$
$\mathrm{sinc}^2(Wk)$	$(1/W)\mathrm{tri}(\omega/2\pi W)(0<W<1/2)$
$\sum_{n=-\infty}^{\infty}\delta[k-nN]$	$(1/N)\sum_{n=-\infty}^{\infty}\delta(\omega/2\pi-n/N)$

例 C-8 离散随机变量的特征函数为：

$$\Phi_X(j\omega)=\mathcal{E}[\exp(j\omega X)]=\sum_{x=-\infty}^{\infty}p_X[x]\exp(j\omega x) \tag{C-45}$$

这是概率质量函数(pmf)$p_X[x]$的 DTFT，不同的是 $\omega\to-\omega$。因此表 C-7 的所有属性都可用于特征函数。需要注意的是表 C-8 中只包括了两个 pmf 的 DTFT，对应于几何和离散均匀随机变量。对于 pmf 为 $p_X[x]=pq^x I_{Z^+}[x]$的几何随机变量(其中 $I_{Z^+}[x]$是离散指示函数)，表中的第四个变换给出了

$$\Phi_X(j\omega)=p/[(1-q\exp(j\omega)] \tag{C-46}$$

其中我们已用 $-\omega$ 代替了 ω。附录 A 中总结了本书中所涉及的离散随机变量的特征函数。◀

进一步阅读

本附录中总结的频域变换可以在 Bracewell(1978)，Lathi(1965)，Lathi(2000)，Mitra(2010)，Oppenheim 和 Schafer(2009)，Oppenheim、Willsky 和 Nawab(1996)，Stearns 和 Hush(1990)，Tretter(1976)中进一步研究。

附录 D

积分法和积分

本附录给出计算随机变量期望的不同类型积分的描述，简要回顾了黎曼积分，这样就可以描述更通用的(不太熟悉的)黎曼-斯蒂尔切斯积分。我们不会像大多数关于测度理论和积分的教科书一样严格，反而对概率空间$\{\mathcal{R}, \mathcal{B}(\mathcal{R}), P_X\}$，提供对以下积分意义的基本理解：

$$\int_{-\infty}^{\infty} g(x)\mathrm{d}F_X(x) \tag{D-1}$$

其中$F_X(x)$是X的累积分布函数(cdf)，$g(x)$是被积函数。我们还简要描述了勒贝格积分，它比黎曼积分更通用，可用于黎曼不可积的函数。对抽象概率空间$\{\Omega, \mathcal{F}, P\}$，它通常写为：

$$\int_{\Omega} g(\omega)\mathrm{d}P(\omega) \tag{D-2}$$

其中$\omega \in \Omega$是样本空间的元素。

这些关于积分的知识对理解第5章中描述的各种矩和期望的定义不是必须的。但它对于理解离散和混合随机变量，以及用于概率密度函数(pdf)和cdf的狄拉克δ函数和阶跃函数的特点有帮助。

D.1 黎曼积分的回顾

黎曼积分通常在微积分课程的开篇介绍，用于第3章和第4章中随机变量的积分。考虑如图D-1所示函数$y=g(x)$，它在$[a,b]$上是连续的和有界的。一般来说一个可积函数是不需要连续的。可微函数形成了连续函数的一个子集，连续函数形成了可积函数的一个子集。后面我们再考虑不连续函数的积分。

例D-1 矩形函数可写为$\mathrm{rect}(x)=u(x+1/2)-u(x-1/2)$，其中$u(x)$是单位阶跃函数。显然$\mathrm{rect}(x)$在$x=\pm 1/2$处是不可微且不连续的，但在$[-1/2,1/2]$上是可积的：它具有单位面积。三角函数$\mathrm{tri}(x)=(1-|x|)\mathrm{rect}(x/2)$在$[-1,-1]$上是连续且可积的，但在$x=0$处是不可微的。 ◀

观察到图D-1中x轴上$[a,b]$被划分成子区间$[x_n,x_{n+1}]$的集合，其中$n=1,\cdots,N$。设α_n为第n个子区间$[x_n,x_{n+1}]$中的一个点。分区的网格是具有最大宽度$\Delta_n \triangleq x_{n+1}-x_n$的子区间(尽管通常子区间具有相同的宽度，如本例所示)。曲线下的面积可由以下黎曼求和近似：

$$R(N)=\sum_{n=1}^{N} g(\alpha_n)(x_{n+1}-x_n)=\sum_{n=1}^{N} g(\alpha_n)\Delta_n \tag{D-3}$$

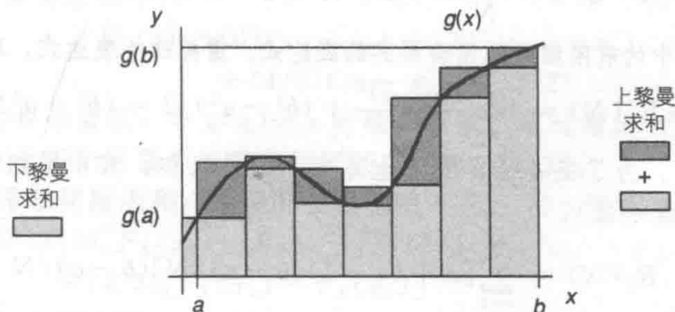


图 D-1 函数$y=g(x)$在x轴上的$[a,b]$分成了子区间。浅阴影矩形组成了下黎曼求和。浅阴影加上深阴影矩形组成上黎曼求和

它是 N 个矩形面积的和；第 n 个矩形宽度为 Δ_n，高度为 $g(\alpha_n)$，面积为 $g(\alpha_n)\Delta_n$。图中显示了两种不同的求和结果，这取决于用函数的哪个值做矩形的高度。下黎曼求和是选 $[x_n, x_{n+1}]$ 上 $g(x)$ 的下确界 m_n 为 α_n 得到的：

$$R_L(N) = \sum_{n=1}^{N} m_n \Delta_n \tag{D-4}$$

于是得到了浅阴影矩形的面积。下黎曼求和的上界必然由 $[a,b]$ 上所有 x 值对应的 $g(x)$ 确定。上黎曼求和则是选 $[x_n, x_{n+1}]$ 上 $g(x)$ 的上确界 M_n 为 α_n 得到的：

$$R_U(N) = \sum_{n=1}^{N} M_n \Delta_n \tag{D-5}$$

这将导致求和值较小的矩形扩展到深色的矩形，这样上黎曼求和的下界由 $[a,b]$ 上所有 x 值对应的 $g(x)$ 确定。因为这些定义中用到了下确界和上确界，x 轴上的子区间可以是半开的，对连续函数可以得到相同的结果。

黎曼积分 R 是通过同时增加间隔数 $N\to\infty$ 和减小它们的宽度 $\Delta_n\to 0$ 导出的。它可以定义如下：对每个 $\delta>0$ 存在 $\varepsilon>0$ 使得

$$\max(\Delta_n) < \varepsilon \Rightarrow \left| \sum_{n=1}^{N} g(\alpha_n)\Delta_n - R \right| < \delta \tag{D-6}$$

而且当 $N\to\infty$ 如果 $R_L(N)\to R_U(N)$，则 $[a,b]$ 上的黎曼积分存在。显然随着矩形宽度的减少和矩形数量的增多，它们的面积和将趋近 $[a,b]$ 上连续函数 $g(x)$ 下的面积。我们将黎曼积分表示如下：

$$R = \int_a^b g(x)\,\mathrm{d}x = G(b) - G(a) \tag{D-7}$$

其中 $G(x)$ 是被积函数 $g(x)$ 的不定积分，则 $G'(x)=g(x)$。最后一个表达式来自微积分基本定理。

在之前求和的讨论中，我们考虑的是非负函数。如果函数有负的部分，则每个正和负的部分的积分要分开计算；从正的部分的结果中减去负的部分的结果得到整个积分：$R=R_+-R_-$。

例 D-2　考虑 $[a,b]$ 上的线性函数 $g(x)=x$，对此函数我们知道

$$\int_a^b x\,\mathrm{d}x = (x^2/2)\big|_a^b = (b^2-a^2)/2 \tag{D-8}$$

对黎曼求和，划分子区间得到 $x_{n+1}=a+n(b-a)/N$ 以及 $\Delta_n=n(b-a)/N-(n-1)(b-a)/N=(b-a)/N$。式(D-3)中选 $\alpha_n=x_{n+1}$（在第 n 个子区间上 $g(x)=x$ 的最大值）得到上黎曼求和：

$$\begin{aligned} R_U(N) &= \sum_{n=1}^{N}[a+n(b-a)/N](b-a)/N = a(b-a)+\frac{(b-a)^2}{N^2}\sum_{n=1}^{N} n \\ &= ab-a^2+\frac{(b-a)^2}{N^2}(N+N^2)/2 \end{aligned} \tag{D-9}$$

上式中代入了附录 E 中的有限级数的闭合形式的表达式。重排这个表达式，取极限得到 R：

$$\lim_{N\to\infty} R_U(N) = \lim_{N\to\infty}\frac{(b-a)^2}{2N}+(b^2-a^2)/2 = (b^2-a^2)/2 \tag{D-10}$$

它是式(D-8)的结果。为了完备性，我们也写出下黎曼求和，其中用到了 $\alpha_n=x_n$（第 n 个子区间中 $g(x)=x$ 的最小值）：

$$R_L(N) = \sum_{n=1}^{N}[a+(n-1)(b-a)/N](b-a)/N \tag{D-11}$$

上式用 $n-1$ 取代了式(D-9)的 n。因此，我们只需将式(D-10)最后的表达式中减去一项即可，如下所示：

$$R_L(N)=\frac{(b-a)^2}{2N}+(b^2-a^2)/2-\frac{(b-a)^2}{N} \tag{D-12}$$

它在 $N\to\infty$ 时也得到 $(b^2-a^2)/2$。 ◀

D.2 黎曼-斯蒂尔切斯积分

黎曼-斯蒂尔切斯积分是黎曼积分的通用化，对式(D-3)中的黎曼求和的子区间 $\{\Delta_n=x_{n+1}-x_n\}$ 做了以下替换：

$$\mathrm{RS}(N)=\sum_{n=1}^{N}g(\alpha_n)[h(x_{n+1})-h(x_n)] \tag{D-13}$$

其中 $h(x)$ 是一个不减函数。通过这种方法，不同的加权值可以分配给在 x 轴上分区的子区间。这种积分是针对实函数 $h(x)$ 进行的；原始函数 $g(x)$ 仍被称为被积函数，$h(x)$ 被看做积分器。黎曼-斯蒂尔切斯积分 RS 用类似于黎曼积分 R 的方式定义如下：对每个 $\delta>0$，存在 $\epsilon>0$ 使得

$$\max(\Delta_n^h)<\epsilon\Rightarrow\left|\sum_{n=1}^{N}g(\alpha_n)\Delta_n^h-\mathrm{RS}\right|<\delta \tag{D-14}$$

其中 $\Delta_n^h\triangleq h(x_{n+1})-h(x_n)$。以下符号用于 $[a,b]$ 上的黎曼-斯蒂尔切斯积分：

$$\mathrm{RS}=\int_a^b g(x)\mathrm{d}h(x) \tag{D-15}$$

如果函数有负的部分，则分别计算每个正的和负的部分的积分，得到总积分结果：$\mathrm{RS}=\mathrm{RS}_+-\mathrm{RS}_-$。

积分器 $h(x)$ 不需要是可微的或连续的；事实上，黎曼-斯蒂尔切斯积分在计算 cdf 不连续的离散和混合随机变量的期望值时非常有用。如果 X 是一个 cdf 为 $F_X(x)$(根据定义它是不减的)的随机变量，则设 $h(x)=F_X(x)$ 得到期望值(第5章中讨论)：

$$\mathcal{E}[g(X)]=\int_a^b g(x)\mathrm{d}F_X(x) \tag{D-16}$$

则我们发现在引言中提到的式(D-1)是由 cdf$F_X(x)$ 加权的 $g(x)$ 的黎曼-斯蒂尔切斯积分。

例 D-3 考虑参数为 p 的伯努利随机变量 X 具有如下的 cdf：

$$F_X(x)=\sum_{n=0}^{1}p^nq^{1-n}u(x-n) \tag{D-17}$$

其中 $u(x)$ 是单位阶跃函数，$q\triangleq 1-p$。假设我们希望计算以下的黎曼-斯蒂尔切斯积分：

$$\mathcal{E}[X]=\int_{-\infty}^{\infty}x\mathrm{d}F_X(x) \tag{D-18}$$

它是 X 的期望值(均值)。图 D-2 给出了被积函数为 $g(x)=x$，积分器为 $h(x)=F_X(x)$，参数 $p=1/2$ 时的曲线图。伯努利 cdf 在 $x=\{0,1\}$ 处有两个左不连续点，但定义为右连续(由实心圆标注了)。x 轴的分区由 $x_1=0$，$x_2=2/5$，$x_3=4/5$ 以及 $x_4=6/5$ 实现。对这种 $N=3$ 的分区，下黎曼求和为：

$$\begin{aligned}R_L(3)&=(0)(x_2-x_1)+(2/5)(x_3-x_2)\\&\quad+(4/5)(x_4-x_3)=12/25\end{aligned} \tag{D-19}$$

它是$[0,6/5]$上的矩形面积，与 $g(x)=x$ 的积分近似。通过增大 N 值，用一种更好的分区，黎曼求和趋近于那个间隔中的真实面积 $(6/5)^2/2=18/25$。 ◀

对下黎曼-斯蒂尔切斯求和，矩形的加权值用 $F_X(x_{n+1})-F_X(x_n)$ 取代了 $x_{n+1}-x_n$：

$$\begin{aligned}\mathrm{RS}_L(3)&=(0)[F_X(x_2)-F_X(x_1)]+(2/5)[F_X(x_3)-F_X(x_2)]\\&\quad+(4/5)[F_X(x_4)-F_X(x_3)]\\&=(0)(1/2-1/2)+(2/5)(1-1)+(4/5)(1-1/2)=2/5\end{aligned} \tag{D-20}$$

其中只有子区间 $[x_3,x_4]$ 对求和做了贡献。显然用更好的分区，只包括 $x=1$ 的子区间对

加权求和有贡献，且在求极限时得到值 $1\times(1-1/2)=1/2$，它是伯努利随机变量的均值：$\mathcal{E}[X]=p=1/2$。

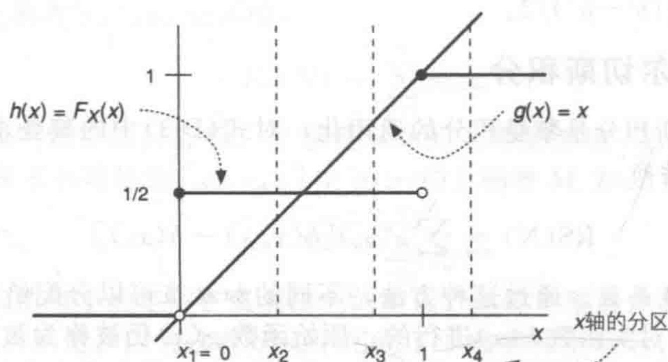


图 D-2 例 D-3 中被积函数为 $g(x)=x$，积分器为 $h(x)=F_X(x)$ 的黎曼-斯蒂尔切斯积分

例 D-4 由例 D-3，我们知道如何计算任意离散随机变量的黎曼-斯蒂尔切斯积分。考虑二项式随机变量：

$$F_X(x)=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}u(x-n) \tag{D-21}$$

当 x 轴的分区随着 N 的增长变得更好后，黎曼-斯蒂尔切斯求和只"抓住了"阶梯函数 $F_X(x)$ 不连续处的 $g(x)$ 的那些值，特别是当 $x_n\to x_{n+1}$ 时，$F_X(x_{n+1})-F_X(x_n)\neq 0$，则对二项式 cdf：

$$\begin{aligned}\mathcal{E}[g(X)]&=\int_{-\infty}^{\infty}g(x)\mathrm{d}F_X(x)=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}\int_{-\infty}^{\infty}g(x)\mathrm{d}u(x-n)\\&=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}g(n)=\sum_{n=0}^{N}p_X[n]g[n]\end{aligned} \tag{D-22}$$

其中 $p_X[n]$ 为 X 的概率质量函数(pmf)。◀

在例 D-3 和例 D-4 中，计算积分(期望)时没有用到狄拉克 δ 函数来表示离散随机变量的 pdf。以前提到过，在有关概率的基本课程中经常用到狄拉克 δ 函数，因为它简化了用黎曼积分计算期望。例如，对二项式随机变量我们可以写为：

$$\mathrm{d}F_X(x)=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}\delta(x-n)\mathrm{d}x \tag{D-23}$$

因此

$$\begin{aligned}\mathcal{E}[g(X)]&=\int_{-\infty}^{\infty}g(x)\mathrm{d}F_x(x)=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}\int_{-\infty}^{\infty}g(x)\delta(x-n)\mathrm{d}x\\&=\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}g(n)=\sum_{n=0}^{N}p_X[n]g(n)\end{aligned} \tag{D-24}$$

其中用到了狄拉克 δ 函数的筛选特性。但是，正如之前的例子所示，黎曼-斯蒂尔切斯积分直接处理了 cdf 的不连续。

如果 $h(x)$ 在 $[a,b]$ 上对通用微分 $h'(x)$ 是处处可微的，则

$$h'(x)=\frac{\mathrm{d}h(x)}{\mathrm{d}x}\Rightarrow\mathrm{d}h(x)=h'(x)\mathrm{d}x \tag{D-25}$$

$$\mathrm{RS}=\int_a^b g(x)h'(x)\mathrm{d}x \tag{D-26}$$

如果 $h(x)=F_X(x)$ 是 cdf，则 $f_X(x)=F_X'(x)$，且

$$\mathrm{RS}=\int_a^b g(x)f_X(x)\mathrm{d}x \tag{D-27}$$

这是已知更通用的求期望的形式。如果 $F_X(x)$ 由连续的 $F_{X,c}(x)$ 和离散的 $F_{X,d}(x)$ 分量混合组成：

$$F_X(x)=F_{X,c}(x)+F_{X,d}(x) \tag{D-28}$$

则期望的计算如下：

$$\mathcal{E}[g(X)]=\int_{-\infty}^{\infty}g(x)\mathrm{d}F_X(x)=\int_{-\infty}^{\infty}g(x)f_{X,c}(x)\mathrm{d}x+\sum_{n=-\infty}^{\infty}g(n)p_{X,d}[n] \tag{D-29}$$

其中 $F_{X,c}(x)$ 是随机变量连续部分的 pdf，$p_{X,d}[n]$ 是离散部分的 pmf。

例 D-5 考虑一个混合随机变量：

$$F_X(x)=(1/2)[1-\exp(-\lambda x)]I_{[0,\infty)}(x)+(1/2)\sum_{n=0}^{\infty}pq^n u(x-n) \tag{D-30}$$

它实际上是一个指数和几何 cdf 的有限混合(见第 4 章)。这个随机变量的均值为：

$$\begin{aligned}\mathcal{E}[X]&=\int_0^{\infty}x\mathrm{d}F_X(x)=(1/2)\int_0^{\infty}x\lambda\exp(-\lambda x)\mathrm{d}x+(1/2)\sum_{n=0}^{\infty}npq^n\\&=(1/2)[1/\lambda+(q/p)]\end{aligned} \tag{D-31}$$

它是指数和几何随机变量的均值的平均。◀

对于有界函数，黎曼和黎曼-斯蒂尔切斯积分的定义非常明确，其中 x 轴可被分成一组非重叠的子区间。回顾之前学过闭集 $[a,b]$ 的长度由区间的勒贝格测度给定：$L([a,b])=b-a$。对不连续函数，比如狄利克函数，定义如下：

$$g(x)\triangleq\begin{cases}1, & x\in\mathcal{Q}\\0, & x\in\mathcal{R}-\mathcal{Q}\end{cases} \tag{D-32}$$

其中 $\mathcal{Q}$ 是有理数集合。在计算这个积分时将域分割成子区间是没有用的。处理这类情况时，我们可将 x 轴上的子区间长度扩展为一个集合的测度(代替子区间的勒贝格测度)。这引出了勒贝格积分。

D.3 勒贝格积分

对函数 $y=g(x)$，特定的 x 值唯一的映射到 y 的一个值，这是函数定义时的性质。如果函数不是一一对应的，那么多个 x 可以映射到相同的 y，这样逆象 $g^{-1}(y)$ 是一对多的。最初描述勒贝格积分时，这种区别是必要的，它可以通过对 y 轴分区来取代对 x 轴分区(如黎曼积分中所做)得到勒贝格积分。对 y 轴分区允许计算黎曼不可积函数的积分，其中 x 轴不能用一个有用方式分区。

考虑图 D-3a 中显示的 $y=g(x)$，其中 y 轴上的间隔 $[g(a),g(b)]$ 被划分成子区间 $\{[y_n,y_{n+1})\}$，这产生了水平的矩形，它们的末端由函数约束。我们要计算对每个 y_n，所有由高度为 y_n 的 x 轴上的子区间所形成的垂直矩形的面积。对于图 D-3b 中的例子，我们看

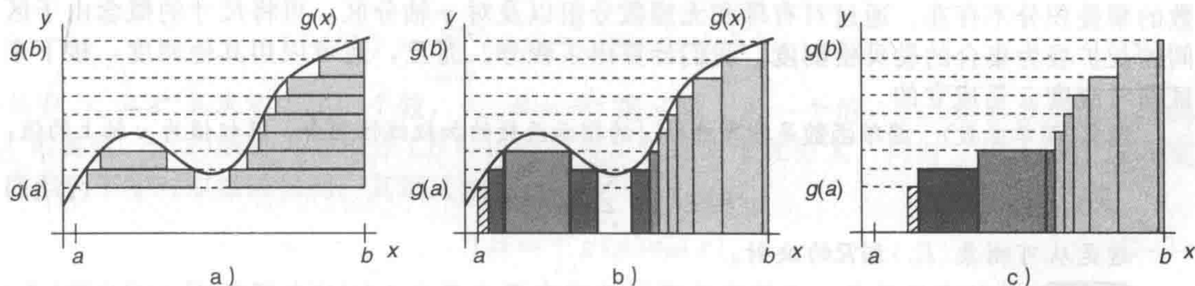


图 D-3 勒贝格求和。a) $[a,b]$ 上的函数 $y=g(x)$ 的 y 轴被分成了子区间 $\{[y_n、y_{n+1})\}$ 产生了水平矩形；b) 由水平矩形端点定义的垂直矩形。那些具有相同阴影的矩形显示的是 x 轴上具有相同矩形高度的子区间；c) 对应于增大的 y 重排的垂直矩形

到有一个高度为$g(a)$(用斜线表示的)的垂直矩形。有三个垂直的矩形(用最深色的阴影表示)，它们的高度由分区中接下来的 y 值给定，然后还有两个矩形，高度为再接下来的 y 值。因为 y 到 x 的逆映射不是一一对应的，所以产生了多个矩形。y 轴分区中其他值对应单一的垂直矩形(最浅色阴影表示的)。如果对 y 轴做更好的分区，则在函数中间“丢失”的垂直矩形会被包括进来。

黎曼积分就是这样做的，通过将矩形面积求和得到勒贝格积分的近似值。勒贝格总和为：

$$\mathrm{LB}(N)=\sum_{n=1}^{N}y_nL(\{x\in[a,b]:y_n\leqslant g(x)<y_{n+1}\}) \tag{D-33}$$

其中 $L(\cdot)$是勒贝格测度，则 $L(\{x\in[a,b]:y_n\leqslant g(x)<y_{n-1}\})$ 给出了所有具有相同高度的矩形的总体宽度。随着 $N\to\infty$，y 轴的分区越来越细化，$\mathrm{LB}(N)$趋近于 $[a,b]$ 上的函数曲线下的面积。显然，如果 $g(x)$是连续和有界的，则黎曼积分和勒贝格积分是相同的：这是显而易见的，因为我们刚刚重排了进行求和近似时的矩形的构造。黎曼求和是相对于 x 轴的分区为不重叠的子区间定义的，而勒贝格对矩形求和则是对 y 轴的分区定义的。

勒贝格积分处理不连续函数时不可能有效地将 x 轴分割成子区间。除了勒贝格测度之外的其他测度，也可用来构造勒贝格求和。但是求和中各项的顺序对积分的最后值是无影响的，因此考虑重排 y 增长的矩形，如图 D-3c 所示。当然，矩形不再遵循函数$g(x)$(在这个图中没有显示)，但累计面积与图 D-3b 中的一样。

现在考虑一个不连续函数 $g(x)$，比如前面提到的狄利克函数。很明显，如果将 x 轴分成子区间，我们不能构建出有意义的矩形来计算 $g(x)$的黎曼积分。然而计算这个函数的勒贝格求和却很容易，因为 y 轴是分区的，$g(x)$只有两个值：$\{0,1\}$。重排 x 轴上数字的顺序，对无理数先分组，然后是有理数，如图 D-3c 所示。对于狄利克函数，在有限区间 $x\in(a,b)$ 上，y 轴的分区只需 $N=2$ 个子区间 $[0,1)$ 和 $[1,\infty)$，如图 D-4 所示。勒贝格求和可写做如下：

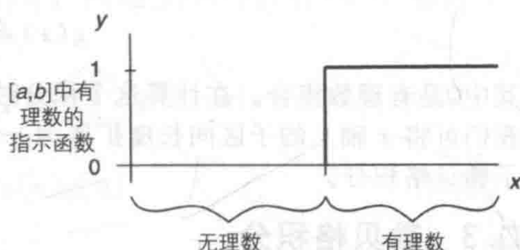


图 D-4 在一个有限区间的实数被重排以使得所有无理数以升序分组，所有有理数也以升序分组的形象例子

$$\begin{aligned}\mathrm{LB}(N)&=\sum_{n=1}^{N}y_nL(\{x\in[a,b]:y_n\leqslant g(x)<y_{n+1}\})\\&=(0)L(\{x\in(a,b]:0\leqslant g(x)<1\})\\&\quad+(1)L(\{x\in[a,b]:1\leqslant g(x)<\infty\})=L(\mathcal{Q})\end{aligned} \tag{D-34}$$

由于 $L(\mathcal{Q})=0$(具有零勒贝格测度的可数集)，狄利克函数的勒贝格积分为零。这个函数的黎曼积分不存在。通过对有理和无理数分组以及对 y 轴分区，再将尺寸的概念由子区间的长扩展为集合的勒贝格测度，我们计算出了积分。而且，还可以用其他测度，接下来证明对测度 μ 是成立的。

定义(简单函数) **简单函数**是指事件$\{E_n\}$的指示函数的加权线性组合，其权值为 y 轴上的值：

$$g(x)=\sum_{n=1}^{N}y_nI_{E_n}(x) \tag{D-35}$$

这是从可测集$\{E_n\}$到$\mathcal{R}$的映射。

例 D-6 狄拉克函数是一个简单函数。只存在两个集合：(i)有理数 $E_1=\mathcal{Q}$；和(ii)无理数 $E_2=\mathcal{R}-\mathcal{Q}$。因此对某个区间 $[a,b]\in\mathcal{R}$：

$$g(x)=y_1I_{\mathcal{Q}}(x)+y_2I_{\mathcal{R}-\mathcal{Q}}(x)=(1)I_{\mathcal{Q}}(x)+(0)I_{\mathcal{R}-\mathcal{Q}}(x)=I_{\mathcal{Q}}(X) \tag{D-36}$$

对所有 $x\in[a,\ b]$都成立。 ◀

需要注意，式(D-35)中的求和由特定的$\{y_n\}$值定义，它可以看作是y轴的分区。没有将x轴分成子区间是因为我们只关注集合$\{E_n\}$和集合中相应的x取值，不考虑x轴的子区间。对于狄拉克函数，这是非常明显的，其中$[a,b]$上的$x\in\mathcal{Q}$是可数的且在此间隔上非均匀分布。类似于式(D-33)，式(D-35)中$g(x)$相对于μ的勒贝格和为

$$\mathrm{LB}(N)=\sum_{n=1}^{N}y_n\mu(E_n) \tag{D-37}$$

勒贝格和相比黎曼和更通用：(Ⅰ)集合$\{E_n\}$被子区间代替；(Ⅱ)子区间的长度被集合的测度μ代替。

式(D-37)的勒贝格和可以扩展到任何如下非负测度函数$g(x)$。考虑将y轴分成等长子区间得到$g(x)$的阶梯近似：

$$g_n(x)=\begin{cases}m/2^n, & m/2^n\leqslant g(x)<(m+1)2^n, 0\leqslant m\leqslant 2^{2n}-1\\ 2^n, & g(x)\geqslant 2^n\end{cases} \tag{D-38}$$

其中m，$n\in\mathcal{Z}$。这种近似方法与信号处理中的量化类似，即用来将连续波形转化为数字信号(在第10章通信中进行了简要讨论)的技术。对于$n=1$，y轴上的子区间为$[0,1/2)$、$[1/2,1)$、$[1,3/2)$、$[3/2,2)$、$[2,\infty)$和相应的$g(x)$的阶梯近似如图D-5所示。由于对y轴划分子区间，并且$g(x)$是x的连续函数，所以x轴也被分成子区间。对于图中的例子，简单函数$g_1(x)$对应的不同y值的个数是5。

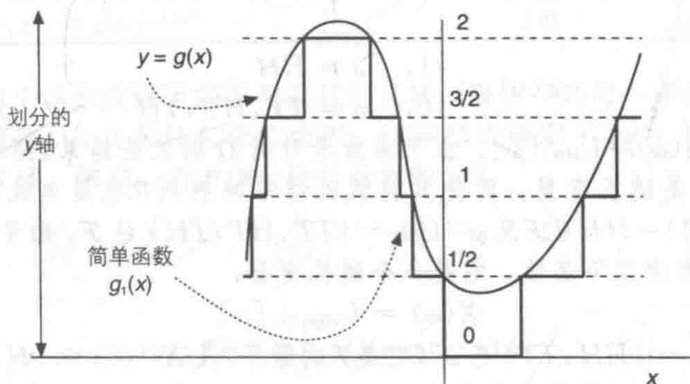


图 D-5 简单函数$g_1(x)$对$g(x)$的阶梯近似

将n增加1，区间的范围随之增大两倍，同时y轴上每个子区间的宽度缩短两倍。由此，通过取极限可以得到

$$\lim_{n\to\infty}\sup|g_n(x)-g(x)|=0 \tag{D-39}$$

其中上确界在所有不超过$g(x)$的简单函数$g_n(x)$之上(图D-5中$n=1$时的例子证明了这一点)。第n个简单函数的勒贝格和为：

$$\mathrm{LB}(N_n)=\sum_{m=0}^{2^{2n}-1}(m/2^n)\mu(E_{m,n})=\sum_{m=0}^{N_n-1}y_{m,n}\mu(E_{m,n}) \tag{D-40}$$

其中$N_n\triangleq 2^{2n}$是求和中项的个数，$y_{m,n}\triangleq m/2^n$是y轴上的一个值，$E_{m,n}$是函数值为$y_{m,n}$的x的集合。$g(x)$的勒贝格积分LB可以通过将步数趋向于无穷大，同时y轴上的子区间宽度趋向于零时取极限得到。其形式如下：

$$\mathrm{LB}=\int_a^b g(x)\mathrm{d}\mu(x) \tag{D-41}$$

如果$g(x)$在$[a,b]$上任意处为负，则可以将其分成正的和负的两部分(如同在黎曼积分中所做的那样)，从正数部分的勒贝格积分减去负数部分的积分：$\mathrm{LB}=\mathrm{LB}_+-\mathrm{LB}_-$。最后，对于抽象概率空间$\{\Omega,\mathcal{F},P\}$，函数$g(\omega)$的期望值为以下勒贝格积分：

$$\text{LB} = \int_{\Omega} g(\omega)\mathrm{d}P(\omega) \tag{D-42}$$

其中 ω 是 Ω 中的元素。由此式(D-2)是 $g(\omega)$ 的勒贝格积分，其测度由概率测度 P 给定，对应于一个抽象样本空间的期望值。

例 D-7 考虑抛两枚公平硬币的试验，其样本空间为 $\Omega=\{HH,TT,HT,TH\}$。令 σ-域 $\mathcal{F}$ 的原子为 $\{\{HH,TT\},TH,HT\}$，由此

$$\mathcal{F}=\{\phi,\Omega,\{HH,TT\},TH,HT,\{TH,HT\},\{HH,TT,TH\},\{HH,TT,HT\}\} \tag{D-43}$$

由于概率测度 P 是为 $\mathcal{F}$ 中的事件定义的，而不是 Ω 的试验结果，除非原子与试验结果相同。所有原子的概率相加必为 1，因为（Ⅰ）它们不相交；（Ⅱ）它们的并集一定为 Ω。由此，对于公平硬币有：

$$\begin{aligned} P(\{HH,TT\}) &= 1/2, P(TH)=P(HT)=1/4 \\ P(\{HH,TT,TH\}) &= P(\{HH,TT,HT\})=3/4 \\ P(\Omega) &= P(\{HH,TT,TH,HT\})=1, P(\phi)=0 \end{aligned} \tag{D-44}$$

对于这种特定的 σ-域，试验结果 $\{HH\}$ 和 $\{TT\}$ 作为事件已经被排除，因此没有单独分配概率；式(D-44)给出了 $\mathcal{F}$ 中所有 8 个事件的概率，这 8 个事件构成幂集 $\mathcal{P}(\Omega)$ 的一个子集。(继续使用幂集的概念，显然 $P(HH)=P(TT)=1/4$。)假设我们只关注以下函数的积分：

$$g(\omega)=\begin{cases}1, & \omega=HH \\ 0, & \omega=TT,HT,TH\end{cases} \tag{D-45}$$

即离散指示函数：$g(\omega)=I_{\{HH\}}[\omega]$。由于函数是针对 Ω 的试验结果(而不是 $\mathcal{F}$ 中的事件)定义的，所以 $g(\omega)$ 不是随机变量。即使其将试验结果映射到 $\mathcal{R}$ 也是如此，因为它是不可测的：其逆像为 $g^{-1}(1)=HH\notin\mathcal{F}$ 及 $g^{-1}(0)=\{TT,HT,TH\}\notin\mathcal{F}$。由于勒贝格积分假设函数是可测的，所以考虑以下函数，它是一个随机变量：

$$X(\omega)=I_{\{HH,TT\}}[\omega] \tag{D-46}$$

注意到 $X^{-1}(1)=\{HH,TT\}\in\mathcal{F}$(它是 $\mathcal{F}$ 的原子)及 $X^{-1}(0)=\{HT,TH\}\in\mathcal{F}$(这是 $\mathcal{F}$ 的两个原子的并)。此时勒贝格积分为：

$$\begin{aligned} \int_{\Omega} X(\omega)\mathrm{d}P(\omega) &= (1)P(\{\omega\in\Omega: X(\omega)=1\})+(0)P(\{\omega\in\Omega: X(\omega)=0\}) \\ &= (1)P(\{HH,TT\})+(0)P(\{HT,TH\}) \\ &= (1)P(\{HH,TT\})+(0)[P(HT)+P(TH)] \\ &= (1)(1/2)+(0)(1/4+1/4)=1/2 \end{aligned} \tag{D-47}$$

我们发现 $X(\omega)$ 是例 D-3 中的伯努利随机变量，其期望值利用黎曼-斯蒂尔切斯积分计算得到。上面的结果即 Ω 在抽象概率空间 $\{\Omega,\mathcal{F},P\}$ 中相应的勒贝格积分，显然与例 D-3 的结果相同。◀

D.4 PDF 的积分

第 3 章以及附录 A 中介绍的各种 pdf 函数对于许多应用的随机试验结果建模非常有用。其中大多数 pdf 包含一个指数函数项，如表 D-1 所示，表中我们将这些指数项分为三个组(除了连续均匀和三角 pdf)。第一组中的 pdf 包含 x^2(或者更高次，例如韦伯分布的情况)的指数函数。第二组中也包含指数函数，但是是 x 的指数函数，而且对数分布和极值分布的 pdf 有两个指数函数项。第三组 pdf 不包含任何指数函数项，以不同的指数依赖于 x。

表 D-1 连续随机变量的 pdf 函数形式

随机变量	函数形式	定义域	Area
Chi	$x^{N-1}\exp(-x^2/2)$	$\mathcal{R}^+$	$\Gamma(N/2)2^{N/2-1}$
高斯	$\exp(-(x-\mu)^2 2\sigma^2)$	$\mathcal{R}$	$\sqrt{2\pi\sigma^2}$
半正态	$\exp(-x^2/2\alpha^2)$	$\mathcal{R}^+$	$\alpha\sqrt{\pi/2}$
反高斯	$(1/x^{3/2})\exp(-\alpha(x-\mu)^2/2\mu^2 x)$	$\mathcal{R}^+$	$\sqrt{2\pi/\alpha}$
对数-正态	$x^{-1}\exp(-(\ln(x)-\mu)^2/2\sigma^2)$	$\mathcal{R}^+$	$\sqrt{2\pi\sigma^2}$
麦克斯韦-玻耳兹曼	$x^2\exp(-x^2/2\alpha^2)$	$\mathcal{R}^+$	$\alpha^3\sqrt{\pi/2}$
瑞利	$x\exp(-x^2/2\alpha^2)$	$\mathcal{R}^+$	α^2
莱斯	$x\exp(-(x^2+c^2)/2\alpha^2)I_0(cx/\alpha^2)$	$\mathcal{R}^+$	α^2
韦伯	$x^{r-1}\exp(-(x/\alpha)^r)$	$\mathcal{R}^+$	α^r/r
卡方	$x^{N/2-1}\exp(-x/2)$	$\mathcal{R}^+$	$\Gamma(N/2)2^{N/2}$
指数	$\exp(-\lambda x)$	$\mathcal{R}^+$	$1/\lambda$
极值	$\exp(-(x-c)/\alpha)\cdot\exp(-\exp(-(x-c)/\alpha))$	$\mathcal{R}$	α
伽马	$x^{r-1}\exp(-\lambda x)$	$\mathcal{R}^+$	$\Gamma(r)/\lambda^r$
拉普拉斯	$\exp(-\lvert x-\mu\rvert/\alpha)$	$\mathcal{R}$	2α
Logistic	$\exp(-\lvert x-\mu\rvert/\alpha)\cdot[1+\exp(-(x-\mu))]^{-2}$	$\mathcal{R}$	α
贝塔	$x^{\alpha-1}(1-x)^{\beta-1}$	$[0,1]$	$B(\alpha,\beta)$
柯西	$[x-c)^2+\alpha^2]^{-1}$	$\mathcal{R}$	π/α
F 分布	$x^{(m-2)/2}/[1+(m/n)x]^{(m+n)/2}$	$\mathcal{R}^+$	$B(m/2,\ n/2)(n/m)^{m/2}$
帕雷托	x^{-r-1}	$[\alpha,\ \infty)$	$1/r\alpha^r$
学生氏 t	$1/(1+x^2/r)^{(r+1)/2}$	$\mathcal{R}$	$\sqrt{r}B(1/2,\ r/2)$

表中还包含每个函数曲线下的面积及其定义域。最后一列每一项的倒数是应用到泛函形式的权重，这样每个 pdf 都具有单位面积。泛函形式确定了 pdf，权重只是对应 pdf 参数特定值的一个常量。例如，考虑伽马随机变量的 pdf

$$f_X(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}\exp(-\lambda x)u(x) \tag{D-48}$$

其中

$$\Gamma(r)\triangleq\int_0^{\infty}v^{r-1}\exp(-v)\mathrm{d}v \tag{D-49}$$

是伽马函数(参阅附录 B)。定义 $\Gamma(r)$的有限积分的被积项类似于 pdf 的泛函形式。由此，$\Gamma(r)$与 Q-函数的角色类似：它是无闭集形式的积分的一个代表。同样对于贝塔函数：

$$B(\alpha,\beta)\triangleq\int_0^1 v^{\alpha-1}(1-v)^{\beta-1}\mathrm{d}v \tag{D-50}$$

其被积函数为表中给出的贝塔随机变量的泛函形式。$B(\alpha,\ \beta)$只是上述有限积分记号，相对 α 和 β 不存在闭集形式的解。

基于这种泛函的视角，可以利用 pdf 面积为 1 来计算被积函数类似于特定 pdf 的泛函形式的有限区间积分。这可以通过重排被积函数(改变变量或者配方)使之类似于已知 pdf 的泛函形式。(附录 E 中将类似的方法用于 pmf 求和。)

例 D-8 考虑以下积分：

$$\int_{-\infty}^{\infty}\exp(-x^2)\mathrm{d}x=(1\sqrt{2})\int_{-\infty}^{\infty}\exp(-y^2/2)\mathrm{d}y \tag{D-51}$$

其积分与高斯 pdf 的泛函形式类似。由此，当 $\mu=0$ 且 $\sigma=1$，可以得到

$$\int_{-\infty}^{\infty}\exp(-x^2)\mathrm{d}x=\sqrt{\pi} \tag{D-52}$$

类似的方法可以用来求得以下定积分：

$$\int_{-\infty}^{\infty}\exp(-(x-\mu)^2/\alpha^2)\mathrm{d}x=\alpha\sqrt{\pi} \tag{D-53}$$

$$\int_{-\infty}^{\infty} \exp(-ax^2+bx+c)\mathrm{d}x = \sqrt{\pi/a}\exp(c+b^2/4a) \tag{D-54}$$

其中我们对指数进行配方得到最后的表达式。◀

例 D-9 假设我们感兴趣的是指数随机变量的矩：

$$\mathcal{E}[X^n] = \int_0^{\infty} \lambda x^n \exp(-\lambda x)\mathrm{d}x \tag{D-55}$$

由表中伽马随机变量的泛函形式，取 $r-1=n$，有

$$\mathcal{E}[X^n] = \lambda\Gamma(n+1)/\lambda^{n+1} = \Gamma(n+1)/\lambda^n \tag{D-56}$$

这正是附录 A 中的表达式。◀

利用这种方法推导定积分的时候必须谨慎，积分的极限必须与表中的定义域一致。例如，表中定义域$\mathcal{R}^+$的麦克斯韦-玻耳兹曼的结果不能直接用于求定义域为$\mathcal{R}$的高斯随机变量的二阶矩。

D.5 不定积分与定积分

对某些分布非常有用的积分公式列举如下：

- 柯西分布：

$$\int \frac{\mathrm{d}x}{a^2+b^2x^2} = (1/ab)\arctan(bx/a) \tag{D-57}$$

$$\int \frac{x\mathrm{d}x}{a^2+x^2} = (1/2)\ln(a^2+x^2) \tag{D-58}$$

$$\int \frac{x^2\mathrm{d}x}{a^2+x^2} = x - a\arctan(x/a) \tag{D-59}$$

- 卡方、指数、伽马、拉普拉斯分布：

$$\int \exp(ax)\mathrm{d}x = (1/a)\exp(ax) \tag{D-60}$$

$$\int x\exp(ax)\mathrm{d}x = (1/a)(x-1/a)\exp(ax) \tag{D-61}$$

$$\int x^2\exp(ax)\mathrm{d}x = (1/a)(x^2-2x/a+2/a^2)\exp(ax) \tag{D-62}$$

$$\int x^m\exp(ax)\mathrm{d}x = (1/a)x^m\exp(ax) - (m/a)\int x^{m-1}\exp(ax)\mathrm{d}x \tag{D-63}$$

$$\int_0^{\infty} x^m\exp(-ax)\mathrm{d}x = \frac{\Gamma(m+1)}{a^{m+1}}, \quad (\text{其中 } a>0, n>-1) \tag{D-64}$$

- Chi、高斯、半正态、麦克斯韦-玻耳兹曼、瑞利分布：

$$\int_0^{\infty} x^{2m}\exp(-ax^2)\mathrm{d}x = \frac{1\times3\times5\cdots(2m-1)}{2^{m+1}a^m}\sqrt{\pi/a} \tag{D-65}$$

$$\int_0^{\infty} x^{2m+1}\exp(-ax^2)\mathrm{d}x = \frac{m!}{2a^{m+1}}, \quad (\text{其中 } a>0) \tag{D-66}$$

- 帕雷托分布：

$$\int \frac{x\mathrm{d}x}{(bx)^m} = -\frac{1/b^2}{(m-2)(bx)^{m-2}}, \quad (m\neq 1,2) \tag{D-67}$$

- 学生氏 t 分布：

$$\int \frac{\mathrm{d}x}{(a^2+x^2)^m} = \frac{1}{2a^2(m-1)}\left[\frac{x}{(a^2+x^2)^{m-1}} + (2m-3)\int\frac{\mathrm{d}x}{(a^2+x^2)^{m-1}}\right] \tag{D-68}$$

$$\int \frac{x\mathrm{d}x}{(a^2+x^2)^{m+1}} = -\frac{1}{2m(a^2+x^2)^m} \tag{D-69}$$

- 韦伯分布：

$$\int_0^{\infty} x^m \exp(-ax^n)\mathrm{d}x = \frac{\Gamma((m+1)/n)}{na^{(m+1)/n}}, \quad (\text{其中 } a>0, m>-1, n>0) \qquad \text{(D-70)}$$

D.6 积分公式

为简单起见，我们给出莱布尼兹积分规则的表达式：

$$\frac{\partial}{\partial v}\int_{a(v)}^{b(v)} g(u,v)\mathrm{d}u = \int_{a(v)}^{b(v)} \frac{\partial}{\partial v}g(u,v)\mathrm{d}u + g(b(v),v)\frac{\partial}{\partial v}b(v) - g(a(v),v)\frac{\partial}{\partial v}a(v) \qquad \text{(D-71)}$$

用分部积分法：

$$\int_{x_1}^{x_2} g(x)\mathrm{d}h(x) = g(x_2)h(x_2) - g(x_1)h(x_1) - \int_{x_1}^{x_2} h(x)\mathrm{d}g(x) \qquad \text{(D-72)}$$

该式采用黎曼-斯蒂尔切斯积分形式。如果函数在 $[x_1,x_2]$ 上处处可微，则 $\mathrm{d}h(x)=h'(x)\mathrm{d}x$ 及 $\mathrm{d}g(x)=g'(x)\mathrm{d}x$，由此得到

$$\int_{x_1}^{x_2} g(x)h'(x)\mathrm{d}x = g(x_2)h(x_2) - g(x_1)h(x_1) - \int_{x_1}^{x_2} h(x)g'(x)\mathrm{d}x \qquad \text{(D-73)}$$

对于不定积分，消除变量 x，式(D-72)可以记作：

$$\int g\mathrm{d}h = gh - \int h\mathrm{d}g \qquad \text{(D-74)}$$

这是微积分课程中非常熟悉的形式。

D.7 特殊函数的双重积分

最后，我们总结一些特定形式的二变量函数的双重积分，包括：$\delta(x-y)$、$u(x-y)$、$\min(x,y)$、$\max(x,y)$ 和 $|x-y|$。其中一些对第 7 章中的均方微积分非常有用，常用于对非平稳随机过程 $X(t)$ 的自相关函数 $R_{XX}(t_1,t_2)$ 求积分。所有的结果依赖于哪个上限较大。

- 狄拉克 δ 函数 $\delta(y-x)$。此函数在 x-y 平面上对于所有 $x=y$ 非零。当我们对 y 求积分时，在直线 $x=y$ 上得到一个单位阶跃函数，图 D-6a 给出了三维阶跃的形状：

$$\int_0^y \delta(y-x)\mathrm{d}y = u(y-x) \qquad \text{(D-75)}$$

  第二重积分为：

$$\int_0^x u(y-x)\mathrm{d}x = \int_0^{\min(x,y)} \mathrm{d}x = \min(x,y) \qquad \text{(D-76)}$$

  如图 D-6b 所示。

- 单位阶跃函数 $u(y-x)$。对 x 轴求积分得到式(D-76)。再对 y 轴积分得到：

$$\int_0^y \min(x,y)\mathrm{d}y = \int_0^x y\mathrm{d}y + \int_x^y x\mathrm{d}y = (yx - x^2/2)u(y-x) \qquad \text{(D-77)}$$

  式中可以理解为 $y\geqslant x$，因为我们从式(D-75)的 $u(y-x)$开始计算。这种二次形式如图 D-6c 所示。其表面并不关于 $y=x$ 对称，因为单位阶跃函数同样不是对称的。当 x 和 y 趋近于 1 时，其最大值为 1/2。$u(y-x)$的双重积分要求交换式(D-77)中的 x 和 y。

- 最小值函数 $\min(x,y)$。对满足 $y\geqslant x$ 的 y 求积分得出式(D-77)。对 x 积分得到：

$$\int_0^x (yx - x^2/2)\mathrm{d}x = (yx^2/2 - x^3/6)u(y-x) \qquad \text{(D-78)}$$

  由此，一般地

$$\int_0^x\int_0^y \min(x,y)\mathrm{d}y\mathrm{d}x = \max(x,y)\min^2(x,y)/2 - \min^3(x,y)/6 \qquad \text{(D-79)}$$

  其图形为 D-7a。对于 $x=y=1$，式(D-79)在图形中为 1/3。

- 最大值函数 $\max(x, y)$。最大值函数的图形如图 D-7b 所示。对 $y\geqslant x$ 的 y 求积分得到：

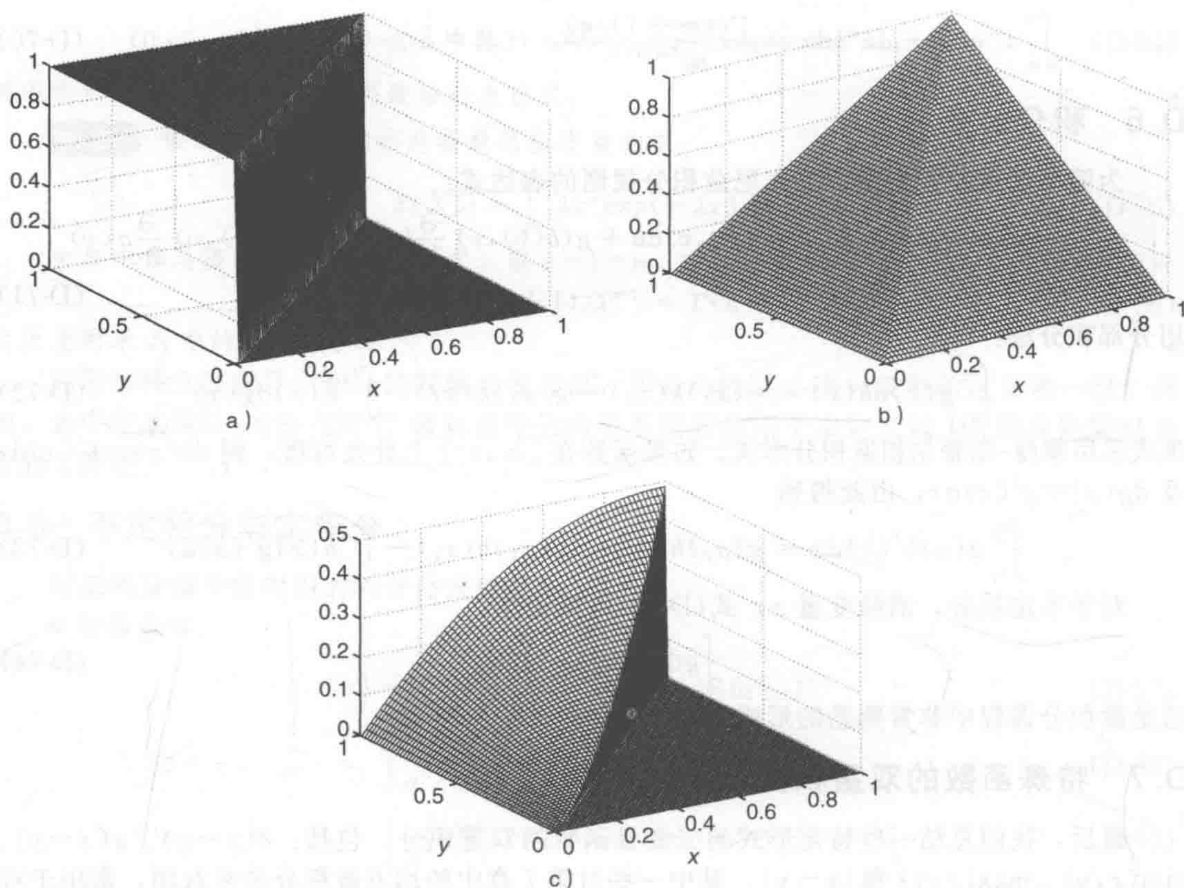


图 D-6　$\delta(y-x)$和$u(y-x)$的积分。a) $\int_0^y \delta(y-x)\mathrm{d}x$；b) $\int_0^x\int_0^y \delta(y-x)\mathrm{d}y\mathrm{d}x$，这也是$\int_0^x u(y-x)\mathrm{d}x$，对应于$\min(x,y)$；c) $\int_0^y\int_0^x u(y-x)\mathrm{d}x\mathrm{d}y$；（注意从 b 到 c 数值范围的变化）

$$\int_0^y \max(x,y)\mathrm{d}y = \int_0^x x\mathrm{d}y + \int_x^y y\mathrm{d}y = (y^2/2 + x^2/2)u(y-x) \tag{D-80}$$

再对 x 求积分得到：

$$\int_0^x (y^2/2 + x^2/2)\mathrm{d}x = (xy^2/2 + x^3/6)u(y-x) \tag{D-81}$$

由此，一般地：

$$\int_0^x\int_0^y \max(x,y)\mathrm{d}y\mathrm{d}x = \min(x,y)\max^2(x,y)/2 + \min^3(x,y)/6 \tag{D-82}$$

此结果如图 D-7c 所示，从中可以看出其增长速度远远超过式(D-79)中 $\min(x,y)$ 的积分。对于 $x=y=1$，式(D-82)的值在图中为 2/3。

- 绝对值函数$|y-x|$。绝对值函数的图形如图 D-8a 所示。对 $y\geqslant x$ 求积分得到：

$$\int_0^y |y-x|\mathrm{d}y = \int_0^x (x-y)\mathrm{d}y + \int_x^y (y-x)\mathrm{d}y = (x^2 + y^2/2 - xy)u(y-x) \tag{D-83}$$

再对 x 求积分得到：

$$\int_0^x (x^2 + y^2/2 - xy)\mathrm{d}x = (x^3/3 + xy^2/2 - x^2y/2)u(y-x) \tag{D-84}$$

由此，一般地：

$$\int_0^x\int_0^y |y-x|\mathrm{d}y\mathrm{d}x = \min^3(x,y)/3 + \min(x,y)\max^2(x,y)/2 - \min^2(x,y)\max(x,y)/2 \tag{D-85}$$

其图形如 D-8b 所示。对于 $x=y=1$，图中式(D-85)的值为 1/3。

图 D-7　$\min(x,y)$ 和 $\max(x,y)$ 的积分。a) $\int_0^x\int_0^y \min(x,y)\mathrm{d}y\mathrm{d}x$；b) $\max(x,\ y)$；c) $\int_0^x\int_0^y \max(x,y)\mathrm{d}y\mathrm{d}x$。(注意从 b 到 c 数值范围的变化)

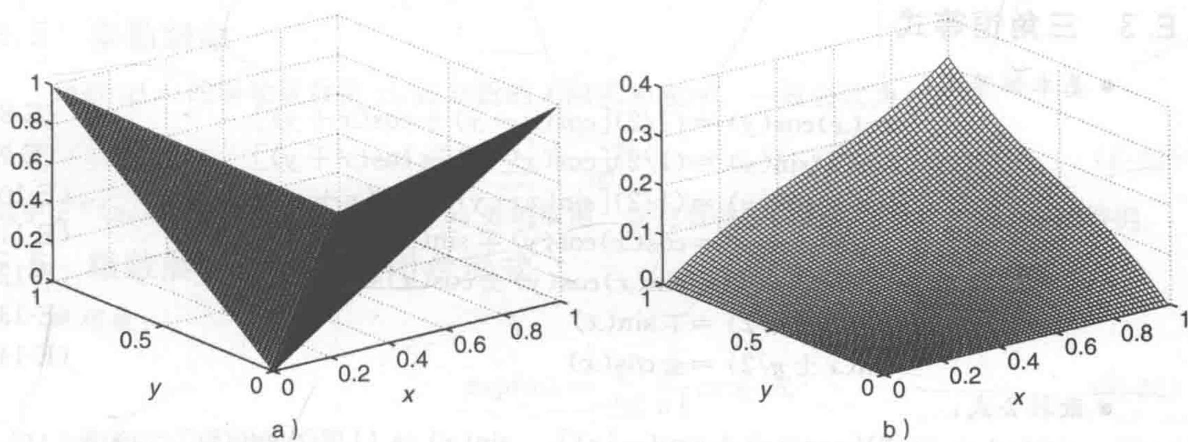


图 D-8　$|y-x|$ 的积分。a) $|y-x|$；b) $\int_0^x\int_0^y |y-x|\mathrm{d}y\mathrm{d}x$；(注意从(a)到(b)数值范围的变化)

进一步阅读

以下参考资料给出不同类型积分的详细信息：Athreya 和 Lahiri(2006)，Billingsley(1986)，Richardson(2009)，Ross(1980)，Sokolnikoff 和 Redheffer(1966)，以及 Thomas，Jr.(1968)。

附录 E

恒等式和无穷序列

该附录包含了前面章节用到的一些恒等式以及对于离散随机变量比较有用的级数。大多数只有结论没有证明。

E.1 零和无穷

- 加减：

$$0+0=0,\quad 0-0=0,\quad 0+\infty=\infty,\quad \infty+\infty=\infty \tag{E-1}$$

$$0-\infty=-\infty,\quad \infty-0=\infty,\quad \infty-\infty=\text{未定义} \tag{E-2}$$

计算柯西分布的均值时会出现式(E-2)中最后一个等式的结果，这解释了为什么其均值不存在。

- 乘除：

$$\infty\cdot\infty=\infty,\quad 0\cdot 0=0,\quad 0\cdot\infty=\text{未定义} \tag{E-3}$$

$$0/0=\text{未定义},\quad \infty/\infty=\text{未定义},\quad 0/\infty=0,\quad \infty/0=\infty \tag{E-4}$$

- 指数：

$$0^0=\text{未定义},\quad 0^\infty=0,\quad \infty^0=\text{未定义},\quad \infty^\infty=\infty \tag{E-5}$$

E.2 最小和最大

当要对包含最小或最大的函数进行积分或微分时，下列恒等式很有用：

$$\min(x_1,x_2)=(1/2)(x_1+x_2-|x_1-x_2|) \tag{E-6}$$

$$\max(x_1,x_2)=(1/2)(x_1+x_2+|x_1-x_2|) \tag{E-7}$$

E.3 三角恒等式

- 基本恒等式：

$$\cos(x)\cos(y)=(1/2)[\cos(x-y)+\cos(x+y)] \tag{E-8}$$

$$\sin(x)\sin(y)=(1/2)[\cos(x-y)-\cos(x+y)] \tag{E-9}$$

$$\sin(x)\cos(y)=(1/2)[\sin(x-y)+\sin(x+y)] \tag{E-10}$$

$$\cos(x\pm y)=\cos(x)\cos(y)\mp\sin(x)\cos(y) \tag{E-11}$$

$$\sin(x\pm y)=\sin(x)\cos(y)\pm\cos(x)\sin(y) \tag{E-12}$$

$$\cos(x\pm\pi/2)=\mp\sin(x) \tag{E-13}$$

$$\sin(x\pm\pi/2)=\pm\cos(x) \tag{E-14}$$

- 欧拉公式：

$$\cos(x)=(1/2)[\exp(\mathrm{j}x)+\exp(-\mathrm{j}x)],\quad \sin(x)=(1/2\mathrm{j})[\exp(\mathrm{j}x)-\exp(-\mathrm{j}x)] \tag{E-15}$$

从而得出欧拉恒等式：$\exp(\mathrm{j}\pi)=-1$。

- 直角和极坐标形式：

$$r\cos(x+\phi)=r\cos(\phi)\cos(x)-r\sin(\phi)\sin(x) \tag{E-16}$$

$$\triangleq a\cos(x)-b\sin(x) \tag{E-17}$$

$$a=r\cos(\phi),\quad b=r\sin(\phi),\quad r=\sqrt{a^2+b^2},\quad \phi=\arctan(b/a) \tag{E-18}$$

- 双曲函数：

$$\sinh(x) = (1/2)[\exp(x) - \exp(-x)] = \sum_{m=0}^{\infty} \frac{x^{2m+1}}{(2m+1)!}$$

$$\cosh(x) = (1/2)[\exp(x) + \exp(-x)] = \sum_{m=0}^{\infty} \frac{x^{2m}}{(2m)!}$$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} \quad \text{(E-19)}$$

$$\cosh^2(x) - \sinh^2(x) = 1, \quad \cosh(x) \pm \sinh(x) = \exp(\pm x) \quad \text{(E-20)}$$

E.4 Stirling公式

阶乘表达式是

$$n! = \prod_{m=0}^{n-1} (n-m) \quad \text{(E-21)}$$

当 n 很大时，可由Stirling公式近似：

$$n! \approx (n/\mathrm{e})^n \sqrt{2\pi n} \quad \text{(E-22)}$$

为方便计算组合问题的概率(包括离散随机变量)，表E-1给出了一些阶乘结果，也给出了由Stirling公式得到的近似值(四舍五入到小数点后一位)。从中看出，即使 n 值相对较小时近似也相当准确。

表E-1 阶乘和Stirling近似

n	$n!$	Stirling近似	n	$n!$	Stirling近似
0	1	0	8	40320	39902.4
1	1	0.9	9	362880	359536.9
2	2	1.9	10	3628800	3598695.6
3	6	5.8	11	39916800	39615625.1
4	24	23.5	12	479001600	475687486.5
5	120	118.0	13	6227020800	6187239475.2
6	720	710.1	14	87178291200	86661001740.6
7	5040	4980.4	15	1307674368000	1300430722199.5

E.5 泰勒级数

函数 $g(x)$ 的泰勒级数由 x_0 处导数的无限项和表示。一般公式为：

$$g(x) = \sum_{n=0}^{\infty} \frac{g^{(n)}(x_0)}{n!}(x-x_0)^n \quad \text{(E-23)}$$

其中 $g^{(n)}(x_0)$ 是 $g(x)$ 的 n 阶导在 $x=x_0$ 处的取值。麦克劳林级数是泰勒级数在 $x_0=0$ 的特例。

E.6 级数展开以及和的闭合形式

- 指数：

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}, x \in \mathcal{R} \quad \text{(E-24)}$$

还给出了有用的结果：

$$\lim_{n\to\infty}(1+x/n)^n = \exp(x) \quad \text{(E-25)}$$

- 三角：

$$\cos(x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n}}{(2n)!}, \quad \sin(x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{(2n+1)!}, x \in \mathcal{R} \quad \text{(E-26)}$$

- 自然对数：

$$\ln(x+1) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)} x^{n+1}, \quad |x| < 1 \quad \text{(E-27)}$$

● 幂和：

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \quad |x|<1 \tag{E-28}$$

$$\sum_{n=m}^{\infty} x^n = \frac{x^m}{1-x}, \quad |x|<1 \tag{E-29}$$

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}, \quad |x|<1 \tag{E-30}$$

$$\sum_{n=1}^{\infty} n^2 x^n = \frac{x(1+x)}{(1-x)^3}, \quad |x|<1 \tag{E-31}$$

$$\sum_{n=1}^{\infty} n^3 x^n = \frac{x(1+4x+x^2)}{(1-x)^4}, \quad |x|<1 \tag{E-32}$$

● 有限项和：

$$\sum_{n=0}^{N} x^n = \frac{1-x^{N+1}}{1-x}, \quad x \neq 1 \tag{E-33}$$

$$\sum_{n=1}^{N} nx^n = \frac{x}{(1-x)^2}[1-(N+1)x^N + Nx^{N+1}], \quad x \neq 1 \tag{E-34}$$

$$\sum_{n=1}^{N} n = (1/2)N(N+1) \tag{E-35}$$

$$\sum_{n=1}^{N} n^2 = (1/6)N(N+1)(2N+1) \tag{E-36}$$

$$\sum_{n=1}^{N} n^3 = (1/4)N^2(N+1)^2 \tag{E-37}$$

$x=1$ 时，可以利用洛必达法则得到式(E-33)以及式(E-34)中闭合形式的结果。可以很容易看出，式(E-33)的和为 $N+1$，式(E-34)的和是式(E-35)。

● 二项式定理：

$$(x+y)^n = \sum_{m=0}^{n} \binom{n}{m} x^{n-m} y^m, \quad x,y \in \mathcal{R}, n \in \mathcal{Z}^+ \tag{E-38}$$

当 $x=y=1$ 时：

$$\sum_{m=1}^{n} \binom{n}{m} = 2^n \tag{E-39}$$

方便起见，表 E-2 提供了一些二项式系数。

表 E-2　二项式系数

n	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$
1	1								
2	2	1							
3	3	3	1						
4	4	6	4	1					
5	5	10	10	5	1				
6	6	15	20	15	6	1			
7	7	21	35	35	21	7	1		
8	8	28	56	70	56	28	8	1	
9	9	36	84	126	126	84	36	9	1
10	10	45	120	210	252	210	120	45	10

● 广义二项式定理：

$$(x+y)^n=\sum_{m=0}^{\infty}\binom{n}{m}x^{n-m}y^m,x,y\in\mathcal{R},|x|>|y|,\quad n\in\mathcal{Z}^+ \tag{E-40}$$

其中上限拓展到了无穷大。以下是$|x|<1$的特殊情况：

$$(1-x)^n=\sum_{m=0}^{\infty}\binom{n}{m}x^m,\quad(1+x)^n=\sum_{m=0}^{\infty}\binom{n}{m}(-x)^m \tag{E-41}$$

● 多项式定理：

$$\left(\sum_{k=1}^{m}x_k\right)^n=\sum_{p_1}\cdots\sum_{p_m}\binom{m}{p_1,\cdots,p_m}x_1^{p_1}\cdots x_m^{p_m} \tag{E-42}$$

从而 $p_1+\cdots+p_m=n$

● 调和级数：

$$\sum_{n=1}^{\infty}\frac{1}{n}=\infty \tag{E-43}$$

这是p级数的特例(发散)。

● p级数：

$$\sum_{n=1}^{\infty}\frac{1}{n^p}=\zeta(p),\quad p>1 \tag{E-44}$$

这是附录B中的泽塔函数。当$p=2$时：

$$\sum_{n=1}^{\infty}\frac{1}{n^2}=\pi^2/6 \tag{E-45}$$

E.7　范德蒙恒等式

定理 E-1(范德蒙)。当 $m,n,k\in\mathcal{Z}^+$ 时：

$$\binom{m+n}{k}=\sum_{p=0}^{k}\binom{m}{p}\binom{n}{k-p} \tag{E-46}$$

证明：考虑以下形式的二项式定理：

$$\sum_{k=0}^{m+n}\binom{m+n}{k}x^k=(x+1)^{m+n} \tag{E-47}$$

这是x的$m+n$次多项式。将上式右侧理解成乘积，并利用二项式定理两次，得到：

$$\sum_{k=0}^{m+n}\binom{m+n}{k}x^k=(x+1)^m(x+1)^n=\sum_{r=0}^{m}\binom{m}{r}x^r\sum_{q=0}^{n}\binom{n}{q}x^q \tag{E-48}$$

将x的指数项组合在一起，重新整理表达式如下：

$$\sum_{r=0}^{m}\binom{m}{r}x^r\sum_{q=0}^{n}\binom{n}{q}x^q=\sum_{k=0}^{m+n}\sum_{p=0}^{k}\binom{m}{p}\binom{n}{k-p}x^k \tag{E-49}$$

因此

$$\sum_{k=0}^{m+n}\binom{m+n}{k}x^k=\sum_{k=0}^{m+n}\sum_{p=0}^{k}\binom{m}{p}\binom{n}{k-p}x^k \tag{E-50}$$

需要注意的当求和下限超过上限时，定义二项式系数为0。k是$m+n$次多项式中x的幂，并且等式两边都关于k求和。由x^k的系数相等得到式(E-46)，从而完成了证明。

例 E-1 利用以下简单情况验证该证明。考虑式(E-48)的乘积$(x+1)^2(x+1)$，其中$m=2$，$n=1$。证明式(E-49)的最后一项是下面的表达式：

$$\sum_{k=0}^{3}\binom{3}{k}x^k=\sum_{k=0}^{3}\sum_{p=0}^{k}\binom{2}{p}\binom{1}{k-p}x^k \tag{E-51}$$

上式左边是：

$$\sum_{k=0}^{3}\binom{3}{k}x^k=\binom{3}{0}x^0+\binom{3}{1}x^1+\binom{3}{2}x^2+\binom{3}{3}x^3=1+3x+3x^2+x^3 \tag{E-52}$$

右边是：

$$\begin{aligned}\sum_{k=0}^{3}\sum_{p=0}^{k}\binom{2}{p}\binom{1}{k-p}x^k=&\binom{2}{0}\binom{1}{0}x^0+\left[\binom{2}{0}\binom{1}{1}+\binom{2}{1}\binom{1}{0}\right]x^1\\&+\left[\binom{2}{0}\binom{1}{2}+\binom{2}{1}\binom{1}{1}+\binom{2}{2}\binom{1}{0}\right]x^2\\&+\left[\binom{2}{0}\binom{1}{3}+\binom{2}{1}\binom{1}{2}+\binom{2}{2}\binom{1}{1}+\binom{2}{3}\binom{1}{0}\right]x^3\end{aligned} \tag{E-53}$$

去掉 4 项等于 0 的二项式系数，得到：

$$\begin{aligned}\sum_{k=0}^{3}\sum_{p=0}^{k}\binom{2}{p}\binom{1}{k-p}x^k=&\binom{2}{0}\binom{1}{0}x^0+\left[\binom{2}{0}\binom{1}{1}+\binom{2}{1}\binom{1}{0}\right]x^1\\&+\left[\binom{2}{1}\binom{1}{1}+\binom{2}{2}\binom{1}{0}\right]x^2+\binom{2}{2}\binom{1}{1}x^3\end{aligned} \tag{E-54}$$

从而

$$\sum_{k=0}^{3}\sum_{p=0}^{k}\binom{2}{p}\binom{1}{k-p}x^k=1+(1+2)x+(2+1)x^2+x^3 \tag{E-55}$$

这与式(E-52)相等。式(E-49)右侧的第一个求和号确定 x 的四个幂次(包括零次)的系数的初始集合，第二个求和号确定这些系数的数目与左侧相匹配。从式(E-48)的右侧注意到：

$$\sum_{r=0}^{2}\binom{2}{r}x^r\sum_{q=0}^{1}\binom{1}{q}x^q=(1+2x+x^2)(1+x)=1+3x+3x^2+x^3 \tag{E-56}$$

这与式(E-55)相等。◀

E.8 概率质量函数和与函数形式

第 3 章中的每个概率质量函数(pmf)要么是有限项和，要么是总和为 1 的无穷级数。大多数随机变量的概率质量函数(不包括离散均匀概率质量函数)的元素具有函数形式，如表 E-3 所示。表中去掉了与 x 无关的常数，最后一列为去掉了这些加权后的函数形式的总和。前五个函数基于成功概率为 p 的伯努利试验。

表 E-3 离散随机变量概率质量函数的函数形式

随机变量	函数形式	取值范围	和
伯努利	$(p/q)^x$	$\{0,1\}$	$1/q$
伯努利(对称)	$(\sqrt{p/q})^x$	$\{-1,1\}$	$1/\sqrt{pq}$
二项式	$\binom{N}{x}(p/q)^x$	$\{0,\cdots,N\}$	$1/q^N$
几何	q^x	$\mathcal{Z}^+$	$1/p$
几何(平移)	q^x	$\mathcal{N}$	q/p
超几何	$\binom{M}{x}\binom{N-M}{n-x}$	$\{0,\cdots,N\}$	$\binom{N}{n}$
对数	$-p^x/x$	$\mathcal{N}$	$\ln(q)$
负二项式	$\binom{x+N-1}{x}q^x$	$\mathcal{Z}^+$	$1/p^N$
泊松	$\alpha^x/x!$	$\mathcal{Z}^+$	$\exp(\alpha)$
Zeta	$1/x^\alpha$	$\mathcal{N}$	$\zeta(\alpha)$

概率质量函数由函数形式确定，表中最后一列的倒数只是函数加权，用以保证概率质量

函数之和为1。例如，对于泊松随机变量，表中的第二列的函数形式是指数函数的级数。因此，通过和的倒数 $1/\exp(\alpha)$ 对函数形式进行加权，以得到有效的概率质量函数。同样的，对于超几何随机变量，函数形式必须由 $1/\binom{N}{n}$ 进行缩放，以获得一个有效的概率质量函数。这些概率质量函数和可以用于得到离散随机变量的各种函数的闭合形式的表达式，通常通过重新排列方程得到可辨识的和。(附录D利用类似的方法得到了概率密度函数的积分)。

E.9　完成配方

考虑二次方程

$$g(x) = ax^2 + bx + c \tag{E-57}$$

我们希望写成以下形式

$$g(x) = a(x + d_1)^2 + d_2 \tag{E-58}$$

很容易看出

$$d_1 = b/2a, \quad d_2 = c - b^2/4a \tag{E-59}$$

提出式(E-57)中的 a，加减 $b^2/4a^2$ 后将式子按如下方式整理证明：

$$\begin{aligned} g(x) &= a[x^2 + (b/a)x] + c \\ &= a[x^2 + (b/a)x + b^2/4a^2 - b^2/4a^2] + c \\ &= a[x^2 + (b/a)x + b^2/4a^2] + c - b^2/4a \end{aligned} \tag{E-60}$$

式中包含所需的平方项：

$$g(x) = a(x + b/2a)^2 + c - b^2/4a \tag{E-61}$$

这种方法经常用于重新排列高斯概率密度函数的指数。

例E-2 当对函数在 x 的整个范围进行积分时，通常可以方便地将积分项写成两部分的乘积：一部分是与 x 有关的常用概率密度函数，另一部分与 x 无关。例如，有可能写成

$$\int_{-\infty}^{\infty} g(x)\mathrm{d}x = \int_{-\infty}^{\infty} c\, f_X(x)\mathrm{d}x = c\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = c \tag{E-62}$$

其中 $f_X(x)$ 是一个有效的概率密度函数，c 是常数或关于 x 以外的其他变量的函数。$f_X(x)$ 的参数也有可能是另一个变量的函数，后面将讨论这种情况。假设我们要对下列函数关于 x 进行积分：

$$g(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x^2 - 2xy)/2\sigma^2) \tag{E-63}$$

得到关于一个变量的函数。完成配方得到

$$\begin{aligned} g(y) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-[(x-y)^2 - y^2]/2\sigma^2)\mathrm{d}x \\ &= \exp(-y^2/2\sigma^2)\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp(-(x-y)^2/2\sigma^2)\mathrm{d}x \end{aligned} \tag{E-64}$$

因为最后一个积分是参数 $\{\mu = y, \sigma\}$ 的高斯概率密度函数，最后的结果为

$$g(y) = \exp(-y^2/2\sigma^2) \tag{E-65}$$ ◀

E.10　分部求和法

分部求和法用来计算序列的有限和，与用于连续函数的分部积分法类似。

引理E-1 对于序列 $x[k]$ 和 $y[k]$：

$$\begin{aligned} \sum_{k=k_1}^{k_2} x[k](y[k+1] - y[k]) &= x[k_2]y[k_2+1] - x[k_1-1]y[k_1] \\ &\quad - \sum_{k=k_1}^{k_2} y[k](x[k] - x[k-1]) \end{aligned} \tag{E-66}$$

证明：将求和移到左侧并将它们合并，去掉 $x[k]y[k]$ 项：

$$\sum_{k=k_1}^{k_2}(x[k]y(k+1]-y[k]x[k-1])=x[k_2]y[k_2+1]-y[k_1]x[k_1-1] \quad \text{(E-67)}$$

消去求和中的相同项，得到右式。

需要注意的是也可能出现其他形式，如

$$\sum_{k=k_1}^{k_2}x[k-1](y[k]-y[k-1])=x[k_2]y[k_2]-x[k_1-1]y[k_1-1]$$

$$-\sum_{k=k_1}^{k_2}y[k](x[k]-x[k-1]) \quad \text{(E-68)}$$

定义差分序列 $\Delta y[k]\triangleq y[k]-y[k-1]$ 和 $\Delta x[k]\triangleq x[k]-x[k-1]$，式(E-68)可写成

$$\sum_{k=k_1}^{k_2}x[k-1]\Delta y[k]=x[k_2]y[k_2]-x[k_1-1]y[k_1-1]-\sum_{k=k_1}^{k_2}y[k]\Delta x[k] \quad \text{(E-69)}$$

该形式类似分部积分：

$$\int_{t_1}^{t_2}x(t)\mathrm{d}y(t)=x(t_2)y(t_2)-x(t_1)y(t_1)-\int_{t_1}^{t_2}y(t)\mathrm{d}x(t) \quad \text{(E-70)}$$

上式利用了黎曼-斯蒂尔切斯积分(见附录D)。

例 E-3 分部求和法可用于验证式(E-34)的有限求和公式。令 $x[k]=k$、$y[k]=x^k$ 得到

$$\sum_{k=1}^{N}k(x^{k+1}-x^k)=(Nx^{N+1}-0\cdot x^1)-\sum_{k=1}^{N}x^k[k-(k-1)]=Nx^{N+1}-\sum_{k=1}^{N}x^k \quad \text{(E-71)}$$

由于 k 不再与 x 相乘，右侧的求和变简单了。左侧提出 $x-1$ 并利用式(E-33)的有限和处理右侧，得到

$$(x-1)\sum_{k=1}^{N}kx^k=Nx^{N+1}-\frac{1-x^{N+1}}{1-x}+1=\frac{x}{1-x}[(N+1)x^N-Nx^{N+1}-1] \quad \text{(E-72)}$$

等式两边同时除以 $x-1$，分子分母同时乘以 -1，得到式(E-34)的结果。◀

进一步阅读

关于本附录中内容的其他信息可以在下列参考文献中找到：Beyer(1976)，Gradshteyn 和 Ryzhik(1980)，Kreyszig(1979)，Sokolnikoff 和 Redheffer(1966)，Weisstein(2003)

附录 F 不等式和期望的界

在这个附录中，我们将介绍为多个期望提供界限的几个不等式。其中一些结论给出了证明过程。

F.1 柯西-施瓦兹和 Hölder 不等式

定理 F-1 （柯西-施瓦兹不等式）。对于随机变量 $\{X,Y\}$ 和函数 $\{g(X),h(Y)\}$：

$$|\mathcal{E}[g(X)h(Y)]| \leqslant \sqrt{\mathcal{E}[g^2(X)]}\sqrt{\mathcal{E}[h^2(Y)]} \tag{F-1}$$

当且仅当 a 为使 $P(g(X)=ah(Y))=1$ 的某些常数时，等号成立。

证明：考虑下列均方误差(Mean-Square Error，MSE)：

$$\mathcal{E}[(g(X)-ah(Y))^2] = \mathcal{E}[g^2(X)] + a^2\,\mathcal{E}(h^2(Y)] - 2a\,\mathcal{E}[g(X)h(Y)] \tag{F-2}$$

其中 $a\in\mathcal{R}$是常数。对该表达式关于 a 求微分，并令结果为0，当

$$a_0 = \frac{\mathcal{E}[g(X)h(Y)]}{\mathcal{E}[h^2(Y)]} \tag{F-3}$$

期望最小。参数 a_0 是该均方误差的标量维纳解(见第 11 章)，乘积 $a_0h(Y)$是给定测量值 $h(Y)$时，$g(X)$的最小均方估计(参见第 9 章)。将 a_0 代入式(F-2)，得到

$$\mathcal{E}[g^2(X)] + \frac{(\mathcal{E}[g(X)h(Y)])^2}{\mathcal{E}[h^2(Y)]} - 2\frac{(\mathcal{E}[g(X)h(Y)])^2}{\mathcal{E}[h^2(Y)]} \geqslant 0 \tag{F-4}$$

和

$$\mathcal{E}[g^2(X)] \geqslant \frac{(\mathcal{E}[g(X)h(Y)])^2}{\mathcal{E}[h^2(Y)]} \tag{F-5}$$

重新整理该表达式，对等式两边开平方，并取绝对值$|\mathcal{E}[g(X)h(Y)]|$，完成了证明。为取等号，令函数成比例 $g(X)=ah(Y)$，其中 a 为实常数。式(F-1)的左侧变为

$$|\mathcal{E}[g(X)h(Y)]| = a\,\mathcal{E}[h^2(Y)] \tag{F-6}$$

右侧是

$$\sqrt{\mathcal{E}[g^2(X)]\,\mathcal{E}[h^2(Y)]} = \sqrt{a^2(\mathcal{E}[h^2(Y)])^2} = a\,\mathcal{E}[h^2(Y)] \tag{F-7}$$

从而等式成立。如果式(F-2)中的均方误差是零，则 $P(g(X)=ah(Y))=1$

例 F-1 考虑函数 $g(X)=(X-\mu_X)/\sigma_X$ 和 $h(Y)=(Y-\mu_Y)/\sigma_Y$，其中 $\{\mu_X,\mu_Y\}$ 和 $\{\sigma_X^2,\sigma_Y^2\}$ 是随机变量 $\{X,Y\}$ 的均值和方差。那么

$$\mathcal{E}[g(X)h(Y)] = \frac{C_{XY}}{\sigma_X\sigma_Y} \triangleq \rho_{XY} \tag{F-8}$$

是第 5 章讨论的相关系数。根据柯西-施瓦兹不等式可以很快得到：

$$|\rho_{XY}| \leqslant \sqrt{\mathcal{E}[g^2(X)]\,\mathcal{E}[h^2(Y)]} = 1 \tag{F-9}$$

柯西-施瓦兹不等式是 Hölder's 不等式的特例。这里我们只给出结论，不提供证明。◀

定理 F-2 (Hölder's 不等式)。对于随机变量 $\{X,Y\}$ 和函数 $\{g(X),h(Y)\}$

$$\mathcal{E}[|g(X)h(Y)|] \leqslant (\mathcal{E}[|g(X)|^p])^{1/p}(\mathcal{E}(|h(Y)|^q])^{1/q} \tag{F-10}$$

其中 p，$q\geqslant1$、$1/p+1/q=1$。当且仅当 a 为使 $P(|g(X)|^p=a|h(Y)|^q)=1$ 的某些常数时，等号成立。

需要注意的是左侧绝对值在期望的内部。当 $p=q=2$ 时

$$\mathcal{E}[|g(X)h(Y)|] \leqslant \sqrt{\mathcal{E}[|g(X)|^2]}\sqrt{\mathcal{E}[|h(Y)|^2]} \tag{F-11}$$

由于$|\mathcal{E}[g(X)h(Y)]|\leqslant\mathcal{E}[|g(X)h(Y)|]$，这是柯西-施瓦兹不等式的另一种形式。

F.2　三角和闵科夫斯基不等式

定理 F-3　（三角不等式）。对于随机变量$\{X,Y\}$：

$$\sqrt{\mathcal{E}[(X+Y)^2]}\leqslant\sqrt{\mathcal{E}[X^2]}+\sqrt{\mathcal{E}[Y^2]} \tag{F-12}$$

证明：将左侧开方里面的式子展开，得到

$$\begin{aligned}\mathcal{E}[(X+Y)^2]&\leqslant\mathcal{E}[|X+Y|(|X|+|Y|)]=\mathcal{E}[|X+Y||X|]+\mathcal{E}[|X+Y||Y|]\\&\leqslant\sqrt{\mathcal{E}[(X+Y)^2]\mathcal{E}[X^2]}+\sqrt{\mathcal{E}[(X+Y)^2]\mathcal{E}[Y^2]}\end{aligned} \tag{F-13}$$

其中右侧两项都利用了柯西-施瓦兹不等式。等式左右两边同除以$\sqrt{\mathcal{E}[(X+Y)^2]}$，完成证明。

三角不等式是闵科夫斯基不等式的特例，这里只给出结论，不提供证明。

定理 F-4　（闵科夫斯基不等式）。对于随机变量$\{X,Y\}$和$p\geqslant1$：

$$(\mathcal{E}[|X+Y|^p])^{1/p}\leqslant(\mathcal{E}[|X|^p])^{1/p}+(\mathcal{E}[|Y|^p])^{1/p} \tag{F-14}$$

当$p=1$时，从式(F-14)可得三角不等式的另一种形式：

$$\mathcal{E}[|X+Y|]\leqslant\mathcal{E}[|X|]+\mathcal{E}[|Y|] \tag{F-15}$$

F.3　比安内梅、切比雪夫和马尔可夫不等式

定理 F-5　（比安内梅不等式）。对于随机变量X和常数c：

$$P(|X-c|\geqslant\delta)\leqslant(1/\delta^n)\,\mathcal{E}[|X-c|^n] \tag{F-16}$$

其中$\delta>0$，$n\in\mathcal{Z}^+$。

证明：定义随机变量$Y\triangleq|X-c|^n$，其概率密度函数为$f_Y(y)$。因为Y和$f_Y(y)$都非负：

$$\mathcal{E}[|X-c|^n]=\int_0^\infty y\,f_Y(y)\mathrm{d}y\geqslant\int_{\delta^n}^\infty yf_Y(y)\mathrm{d}y \tag{F-17}$$

其中下限从零增加。取被积函数中y的最小值并把它从积分中提出来，得到

$$\mathcal{E}[|X-c|^n]\geqslant\delta^n\int_{\delta^n}^\infty f_Y(y)\mathrm{d}y=\delta^nP(|X-c|^n\geqslant\delta^n)=\delta^nP(|X-c|\geqslant\delta) \tag{F-18}$$

重新整理最后一个式子，完成证明。

定理 F-6　（切比雪夫不等式）。对于均值为μ_X、方差为σ_X^2的随机变量X，在任意$\delta>0$时，有：

$$P(|X-\mu_X|\geqslant\delta)\leqslant\sigma_X^2/\delta^2 \tag{F-19}$$

证明：切比雪夫不等式是比安内梅不等式在$n=2$、$c=\mu_X$的一个特例。

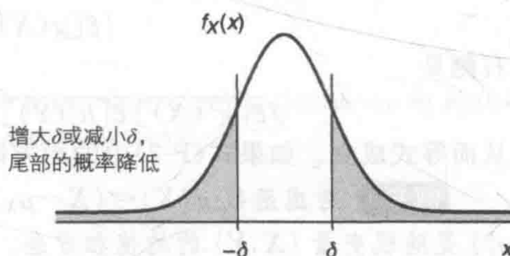


图 F-1　切比雪夫不等式与分布的尾部偏离均值距离的概率有关，这里给出的是参数$\{\mu=0,\sigma=\sqrt{2}\}$的高斯随机变量

该不等式说明了X的“绝大多数”值接近潜在概率密度函数的均值。结果很直观，因为（ⅰ）随机变量偏离均值某些距离的概率；（ⅱ）概率密度函数的宽度（方差）之间存在联系。图 F-1 借由高斯随机变量说明了这一联系。

定理 F-7　对于非负函数$g(x)$，在任意$\delta>0$时，比安内梅不等式有以下一般形式

$$P(g(X)\geqslant\delta)\leqslant\mathcal{E}[g(X)]/\delta \tag{F-20}$$

证明：函数的期望：

$$\mathcal{E}[g(X)]=\int_{-\infty}^\infty g(x)f_X(x)\mathrm{d}x=\int_{-\infty}^{g(x)<\delta}g(x)f_X(x)\mathrm{d}x+\int_{g(x)\geqslant\delta}^\infty g(x)f_X(x)\mathrm{d}x \tag{F-21}$$

因为$g(x)$和$f_X(x)$都非负，因此这两个积分都非负。从而

$$\mathcal{E}[g(X)] \geqslant \int_{g(x)\geqslant\delta}^{\infty} g(x) f_X(x)\mathrm{d}x \geqslant \int_{g(x)\geqslant\delta}^{\infty} \delta\, f_X(x)\mathrm{d}x = \delta P(g(X) \geqslant \delta) \quad \text{(F-22)}$$

其中 $g(x)$已被替换其最小值 δ。重新整理最后一个表达式，完成证明。

特别要注意，前面的不等式中 $g(x)$必须是非负函数。

定理 F-8 （马尔可夫不等式）。对于随机变量 X：

$$P(|X| \geqslant \delta) \leqslant \mathcal{E}[|X|]/\delta \quad \text{(F-23)}$$

对于非负随机变量 Y：

$$P(Y \geqslant \delta) \leqslant \mathcal{E}[Y]/\delta \quad \text{(F-24)}$$

两个不等式对于任意 $\delta>0$ 都适用。

证明：马尔可夫不等式是比安内梅不等式在 $n=1$、$c=0$ 时的特例。

马尔可夫不等式的一般表达式是：对于 $n\in\mathcal{Z}^+$

$$P(|X| \geqslant \delta) \leqslant \mathcal{E}[|X|^n]/\delta^n \quad \text{(F-25)}$$

F.4 切尔诺夫不等式

定理 F-9 （切尔诺夫不等式）。对于随机变量 X 和任意 $\delta>0$：

$$P(X \geqslant \delta) \leqslant \exp(-t\delta)\ \mathcal{E}[\exp(tX)] \quad \text{(F-26)}$$

其中，$\mathcal{E}[\exp(tX)] \triangleq m_X(t)$是 X 和 $t\geqslant 0$ 的矩母函数。

证明：根据矩母函数的定义：

$$m_X(t) = \int_{-\infty}^{\delta} \exp(tx) f_X(x)\mathrm{d}x + \int_{\delta}^{\infty} \exp(tx) f_X(x)\mathrm{d}x \geqslant \int_{\delta}^{\infty} \exp(tx) f_X(x)\mathrm{d}x$$

$$\geqslant \exp(t\delta)\int_{\delta}^{\infty} f_X(x)\mathrm{d}x = \exp(t\delta)P(X \geqslant \delta) \quad \text{(F-27)}$$

由于 $\exp(tx)\geqslant 0$，得到了第一个不等式；将被积函数中的 $\exp(tx)$替换成最小值 $\exp(t\delta)$，得到第二个不等式。重新整理最后一个表达式，完成证明。

由于切尔诺夫不等式需要满足所有的 $t\geqslant 0$，最严格的约束是：

$$P(X \geqslant \delta) \leqslant \min_{t\geqslant 0}(\exp(-t\delta)m_X(t)) = \min_{t\geqslant 0}(\exp(-t\delta + \ln(m_X(t)))) \quad \text{(F-28)}$$

这依赖于 δ 和特定矩母函数的参数。

例 F-2 考虑参数为 λ 的指数随机变量

$$P(X \geqslant \delta) = \lambda\int_{\delta}^{\infty} \exp(-\lambda x)\mathrm{d}x = \exp(-\lambda\delta) \quad \text{(F-29)}$$

由于 X 非负，$\mathcal{E}[X]=1/\lambda$ 时，马尔可夫不等式变为

$$P(X \geqslant \delta) \leqslant 1/\lambda\delta \quad \text{(F-30)}$$

切尔诺夫不等式为

$$P(X \geqslant \delta) \leqslant \min_{t\geqslant 0}(\exp(-t\delta + \ln(\lambda) - \ln(\lambda - t))) \quad \text{(F-31)}$$

其中指数随机变量的 $m_X(t)=\lambda/(\lambda-t)$。对表达式的指数部分关于 t 进行微分，令结果为 0 得到“最优”t_0：

$$-\delta + 1/(\lambda - t_0) = 0 \Rightarrow t_0 = \lambda - 1/\delta \quad \text{(F-32)}$$

注意只有当 $\delta\geqslant 1/\lambda$ 时，$t_0\geqslant 0$；当 $\delta<1/\lambda$ 时，$t_0=0$。◀

将这些结果代入切尔诺夫不等式，得到

$$P(X \geqslant \delta) \leqslant \begin{cases} \lambda\delta\exp(1-\lambda\delta), & \delta \geqslant 1/\lambda \\ 1, & 0 \leqslant \delta < 1/\lambda \end{cases} \quad \text{(F-33)}$$

$\lambda=1$ 时，边界和准确概率与 δ 的关系如图 F-2a 所示。观察发现马尔可夫限并不严格小于 1，当 δ 较大时，切尔诺夫限更严格。标准高斯随机变量的结果如图 F-2b 所示。相应的式子是(用切比雪夫代替马尔可夫)

$$P(|X| \geqslant \delta) = 2Q(\delta) \quad \text{（准确）} \quad \text{(F-34)}$$

$$P(|X|\geqslant\delta)\leqslant 1/\delta^2 \quad (\text{切比雪夫}) \tag{F-35}$$

$$P(|X|\geqslant\delta)\leqslant 2\min_{t\geqslant 0}(\exp(-t\delta+t^2/2)) \quad (\text{切尔诺夫}) \tag{F-36}$$

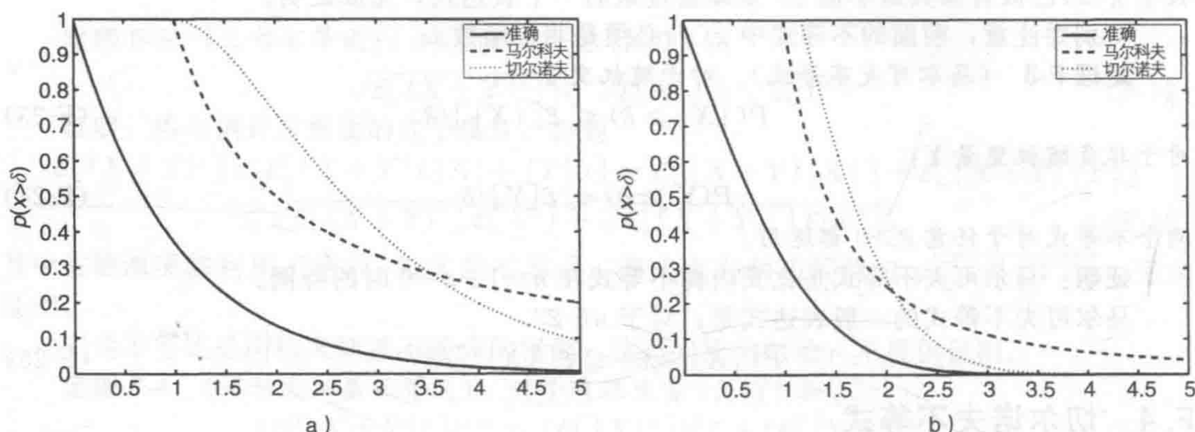


图 F-2　例 F-2 的概率界。a)$\lambda=1$ 的指数分布；b)$\{\mu=0, \sigma=1\}$的标准高斯分布

其中 $Q(\cdot)$是 Q 函数。需要注意的是准确结果和切尔诺夫限都包含了因子 2，这是因为概率涉及$|X|$，该对称分布的两个尾部都要考虑。当 $t=\delta$ 时，最后一个式子取最小，因此图中利用的切尔诺夫限是

$$P(|X|\geqslant\delta)\leqslant 2\exp(-\delta^2/2) \tag{F-37}$$

F.5　Jensen 不等式

定理 F-10　(Jensen 不等式)。设 X 是定义在开区间的随机变量，其累积分布函数为 $F_X(x)$，$g(x)$是一个凸函数。则

$$g(\mathcal{E}[X])\leqslant\mathcal{E}[g(X)] \tag{F-38}$$

证明： 利用凸函数的定义(见附录 B)，存在一条线 $c(x)=a(x-x_0)+g(x_0)$，使得对于开区间中的所有 x，都有 $c(x)\leqslant g(x)$。因此对于随机变量 X：

$$\mathcal{E}[c(X)]\leqslant\mathcal{E}[g(X)] \tag{F-39}$$

令 $c(X)$中的 $x_0=\mathcal{E}[X]$，则

$$\mathcal{E}[c(X)]=a(\mathcal{E}[X]-\mathcal{E}[X])+g(\mathcal{E}[X])=g(\mathcal{E}[X]) \tag{F-40}$$

将该结果代入式(F-39)，完成证明。如果 $g(x)$是严格凸函数，则式(F-38)是严格不等，除非 $g(X)=\mathcal{E}[g(X)]$的概率是 1。

反过来，如果 $g(x)$是一个凹函数，式(F-38)的不等式变号。下面的例子给出了信息论(参见第 10 章)的结果，对于第 9 章的最大期望算法同样有用。

例 F-3　考虑概率密度函数为 $f_X(x)$的随机变量 X，设 $g(x)$是一个非负函数(可以是也可以不是概率密度函数)。定义随机变量 $Y\triangleq g(X)/f_X(X)$。虽然 $f_X(x)$是 X 的概率密度函数，这里用它作为随机变量 X 的函数。第 5 章中提到计算$\mathcal{E}[Y]$不一定要找到 $f_Y(y)$；相反，

$$\mathcal{E}[Y]=\int_{-\infty}^{\infty}y(x)f_X(x)\mathrm{d}x \tag{F-41}$$

代入 $y(x)=g(x)/f_X(x)$，式子简化为

$$\mathcal{E}[g(X)/f_X(X)]=\int_{-\infty}^{\infty}g(x)\mathrm{d}x \tag{F-42}$$

这是 $g(x)$下的面积。接下来考虑凸函数 $\log(Y)$，得到

$$\mathcal{E}[\log(g(X)/f_X(X))]=\int_{-\infty}^{\infty}\log(g(x)/f_X(x))f_X(x)\mathrm{d}x \tag{F-43}$$

由 Jensen 不等式得到下限：

$$\mathcal{E}[\log(g(X)/f_X(X))] \geqslant \log(\mathcal{E}[g(X)/f_X(X)]) = \log\left(\int_{-\infty}^{\infty} g(x)\mathrm{d}x\right) \tag{F-44}$$

其中利用了(F-42)。如果 $g(x)$ 恰好是概率密度函数，则式(F-42)等于1，下限变为

$$\mathcal{E}[\log(g(X)/f_X(X))] \geqslant 0 \tag{F-45}$$ ◀

F.6 克拉美罗不等式

克拉美罗不等式提供了估计方差的下限，它也称为克拉美罗下限(Cramer-Rao lower bound，CRLB)。CRLB可以作为分布参数的有偏和无偏估计性能的度量，如第9章所述。这里提供了一个无偏估计的推导，并给出了有偏估计以及参数向量估计的结果。

定理 F-11 （克拉美罗不等式）。设 T 是标量参数 θ 的函数 $g(\theta)$ 的一个无偏估计，其中 $\{X_1,\cdots,X_N\}$ 是概率密度函数 $f_X(x;\theta)$ 的独立同分布的样本。令 $T=t$ 是样本 $\{x_1,\cdots,x_N\}$ 的特定估计值，并满足下列正则条件：

- 对参数求导

$$\frac{\partial}{\partial\theta}\log(f_X(x;\theta)) \text{ 对于任意 } x \text{ 和 } \theta \text{ 都存在} \tag{F-46}$$

- 积分和求导交换Ⅰ：

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n \tag{F-47}$$

- 积分和求导交换Ⅱ：

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{n=1}^{N}t\,f(x_n;\theta)\mathrm{d}x_n = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}t\frac{\partial}{\partial\theta}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n \tag{F-48}$$

- 正期望的取值范围：

$$0 < \mathcal{E}\left[\left(\frac{\partial}{\partial\theta}\log(f_X(X;\theta))\right)^2\right] < \infty \tag{F-49}$$

θ 的无偏估计 T 的方差下限为

$$\mathrm{var}[T] \geqslant \frac{\left(\frac{\partial}{\partial\theta}g(\theta)\right)^2}{N\,\mathcal{E}\left[\left(\frac{\partial}{\partial\theta}\log(f_X(X;\theta))\right)^2\right]} \tag{F-50}$$

需要注意的是随机变量 X 出现在概率密度函数的参数中。

证明： 由于 T 是 $g(\theta)$ 的无偏估计，

$$g(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}t\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n \tag{F-51}$$

求导得到

$$\frac{\partial}{\partial\theta}g(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}t\frac{\partial}{\partial\theta}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n \tag{F-52}$$

由于

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n = 1 \tag{F-53}$$

式(F-52)可以扩写成：

$$\frac{\partial}{\partial\theta}g(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}t\frac{\partial}{\partial\theta}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n - g(\theta)\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n \tag{F-54}$$

上式只是减去了0，因为它是对式(F-53)中的常数求导。从而

$$\frac{\partial}{\partial\theta}g(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[t-g(\theta)]\frac{\partial}{\partial\theta}\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n \tag{F-55}$$

对于任意函数 $c(\theta)$，都有以下等式：

$$\frac{\partial}{\partial\theta}\log(c(\theta)) = \frac{1}{c(\theta)}\frac{\partial}{\partial\theta}c(\theta) \Rightarrow \frac{\partial}{\partial\theta}c(\theta) = c(\theta)\frac{\partial}{\partial\theta}\log(c(\theta)) \tag{F-56}$$

在式(F-55)中利用这一性质，得到

$$\frac{\partial}{\partial\theta}g(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[t-g(\theta)]\frac{\partial}{\partial\theta}\log\Big(\prod_{n=1}^{N}f(x_n;\theta)\Big)\prod_{n=1}^{N}f(x_n;\theta)\mathrm{d}x_n$$

$$= \mathcal{E}\Big[[T-g(\theta)]\frac{\partial}{\partial\theta}\log\Big(\prod_{k=n}^{N}f(X_n;\theta)\Big)\Big] \tag{F-57}$$

从柯西–施瓦兹不等式：

$$\Big(\frac{\partial}{\partial\theta}g(\theta)\Big)^2 \leqslant \mathcal{E}[(T-g(\theta))^2]\,\mathcal{E}\Big[\Big(\frac{\partial}{\partial\theta}\log\Big(\prod_{n=1}^{N}f(X_n;\theta)\Big)\Big)^2\Big] \tag{F-58}$$

其中等式右边第一个期望是 $g(\theta)$ 的无偏估计 T 的方差。由于样本 $\{X_1,\cdots,X_N\}$ 独立同分布：

$$\mathcal{E}\Big[\Big(\frac{\partial}{\partial\theta}\log\Big(\prod_{n=1}^{N}f(X_n;\theta)\Big)\Big)^2\Big] = N\,\mathcal{E}\Big[\Big(\frac{\partial}{\partial\theta}\log(f(X;\theta))\Big)^2\Big] \tag{F-59}$$

将该式代入式(F-58)并重新排列该结果，完成了证明。

重点要注意 CRLB 不依赖于特定无偏估计值 T。

定义(Fisher 信息) 式(F-50)中的分母称为 Fisher 信息：

$$\mathcal{I}(\theta) \triangleq N\,\mathcal{E}\Big[\Big(\frac{\partial}{\partial\theta}\log(f_X(X;\theta))\Big)^2\Big] \tag{F-60}$$

上述对数未指定底，但最常见的是以 e 为底。

定理 F-12 如果将上述正则性条件拓展到二阶导数，Fisher 信息也可表示如下：

$$\mathcal{I}(\theta) = -N\,\mathcal{E}\Big[\frac{\partial^2}{\partial\theta^2}\log(f_X(X;\theta))\Big] \tag{F-61}$$

证明：考虑下列导数：

$$N\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(x;\theta)\mathrm{d}x = 0 \tag{F-62}$$

由于对有效概率密度函数积分的结果求导，上式等于零。交换积分和求导的顺序，利用式(F-56)并令 $c(\theta)=f(x;\ \theta)$，得到

$$N\,\mathcal{E}\Big[\frac{\partial}{\partial\theta}\log(f(X;\theta))\Big] = 0 \tag{F-63}$$

再求导并使用链式法则，得到

$$N\,\mathcal{E}\Big[\frac{\partial^2}{\partial\theta^2}\log(f(X;\theta))\Big] + N\,\mathcal{E}\Big[\Big(\frac{\partial}{\partial\theta}\log(f(X;\theta))\Big)^2\Big] = 0 \tag{F-64}$$

第二项是式(F-60)中的 Fisher 信息，从而完成了证明。

例 F-4 对于参数为 λ 的指数随机变量，将 N 个样本的均值作为 $g(\lambda)=1/\lambda$ 的估计。Fisher 信息为

$$\mathcal{I}(\lambda) = -N\,\mathcal{E}\Big[\frac{\partial^2}{\partial\lambda^2}\ln(\lambda\exp(-\lambda X))\Big] = N/\lambda^2 \tag{F-65}$$

而且由于

$$\frac{\partial}{\partial\lambda}g(\lambda) = -1/\lambda^2 \tag{F-66}$$

该 CRLB 是

$$\mathrm{var}[T] \geqslant \frac{1}{N\lambda^2} \tag{F-67}$$

代入$\mathcal{E}[\overline{X}]=1/\lambda$，样本均值的方差为

$$\mathcal{E}[(\overline{X}-1/\lambda)^2]=\mathcal{E}[\overline{X}^2]-(2/\lambda)\,\mathcal{E}[\overline{X}]+1/\lambda^2=\mathcal{E}[\overline{X}^2]-1/\lambda^2 \tag{F-68}$$

由于样本独立同分布

$$\begin{aligned}\mathcal{E}[\overline{X}^2]&=\frac{1}{N^2}(N\,\mathcal{E}[X^2]+(N^2-N)(\mathcal{E}[X])^2)\\&=\frac{1}{N}[2/\lambda^2+(N-1)/\lambda^2]=\frac{N+1}{N\lambda^2}\end{aligned} \tag{F-69}$$

将这个结果与式(F-68)结合，我们发现样本均值达到了式(F-67)中的CRLB；$T=\overline{X}$ 是 $g(\lambda)$ 的一致最小方差无偏估计(UMVU)(见第9章)。◀

定理 F-13 对于偏差为 $B(\theta)$ 的 $g(\theta)$ 的有偏估计 T

$$\mathcal{E}[T]=g(\theta)+B(\theta) \tag{F-70}$$

CRLB 是

$$\operatorname{var}[T]\geqslant\frac{\left[\dfrac{\partial}{\partial\theta}(g(\theta)+B(\theta))\right]^2}{N\,\mathcal{E}\left[\left(\dfrac{\partial}{\partial\theta}\log(f_X(X;\theta)\right)^2\right]} \tag{F-71}$$

证明：参见习题9-35。

最后，考虑有 M 个参数的向量 $\boldsymbol{\theta}\triangleq[\theta_1,\cdots,\theta_M]^{\mathrm{T}}$，并令 $\boldsymbol{g}(\boldsymbol{\theta})\triangleq[g_1(\boldsymbol{\theta}),\cdots,g_K(\boldsymbol{\theta})]^{\mathrm{T}}$ 是参数函数的向量。将 $\boldsymbol{g}(\boldsymbol{\theta})$ 的 k 个元素的估计定义成向量 $\boldsymbol{T}\triangleq[T_1,\cdots,T_K]^{\mathrm{T}}$，并假定它无偏：$\mathcal{E}[\boldsymbol{T}]=\boldsymbol{g}(\boldsymbol{\theta})$。令 $\boldsymbol{J}\in\mathcal{R}^{K\times M}$ 是下列雅可比矩阵(见附录G)：

$$\boldsymbol{J}\triangleq\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{g}(\boldsymbol{\theta})=\begin{bmatrix}\dfrac{\partial}{\partial\theta_1}g_1(\boldsymbol{\theta}) & \cdots & \dfrac{\partial}{\partial\theta_M}g_1(\boldsymbol{\theta})\\ \vdots & & \vdots\\ \dfrac{\partial}{\partial\theta_1}g_K(\boldsymbol{\theta}) & \cdots & \dfrac{\partial}{\partial\theta_M}g_K(\boldsymbol{\theta})\end{bmatrix} \tag{F-72}$$

定理 F-14 对于参数向量 $\boldsymbol{g}(\boldsymbol{\theta})$ 的无偏估计 $\boldsymbol{T}$：

$$\operatorname{cov}[\boldsymbol{T}]\geqslant\boldsymbol{J}\,\mathcal{I}^{-1}(\boldsymbol{\theta})\boldsymbol{J}^{\mathrm{T}} \tag{F-73}$$

其中

$$\mathcal{I}(\boldsymbol{\theta})=\mathcal{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\log(f(X;\boldsymbol{\theta}))\left(\frac{\partial}{\partial\boldsymbol{\theta}}\log(f(X;\boldsymbol{\theta}))\right)^{\mathrm{T}}\right] \tag{F-74}$$

$$=-\mathcal{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\log(f(X;\boldsymbol{\theta}))\right] \tag{F-75}$$

是 Fisher 信息矩阵 $\in\mathcal{R}^{M\times M}$

式(F-73)在半正定(PSD)范畴定义，这意味着对于任意非零向量 $\boldsymbol{a}$，

$$\boldsymbol{a}^{\mathrm{T}}\operatorname{cov}[\boldsymbol{T}]\boldsymbol{a}\geqslant\boldsymbol{a}^{\mathrm{T}}\boldsymbol{J}\,\mathcal{I}^{-1}(\boldsymbol{\theta})\boldsymbol{J}^{\mathrm{T}}\boldsymbol{a} \tag{F-76}$$

或者说，式(F-73)的左侧部分减去右侧部分得到的矩阵是半正定的。

进一步阅读

本附录中的不等式在第5、9、10章末尾的参考文献中有所讨论。包括 Lehmann(1983)，McEliece(1977)，Mendel(1995)，Mood、Graybill、Boes(1974)，Papoulis(1965)。

附录 G

矩阵和向量的性质

最后的附录中，总结了矩阵和向量的一些重要性质。主要关注元素是实数的矩阵和向量，但也提供了一些复矩阵和向量的结果。

G.1 基本性质

考虑下列 M 行$\{\tilde{\boldsymbol{a}}_m^{\mathrm{T}}\}$ N 列$\{\boldsymbol{a}_n\}$的矩阵 $\boldsymbol{A}\in\mathcal{R}^{M\times N}$：

$$\boldsymbol{A}=[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_N]=\begin{bmatrix}\tilde{\boldsymbol{a}}_1^{\mathrm{T}}\\ \vdots \\ \tilde{\boldsymbol{a}}_M^{\mathrm{T}}\end{bmatrix} \tag{G-1}$$

其中，上标^{T}表示矩阵/向量的转置。$\boldsymbol{A}$ 的标量元素由$\{a_{mn}\}$表示。本书中的所有向量都定义成列向量；行向量通过使用 T 得到。

定义(线性相关) 如果存在非零$\{x_n\}$满足下列条件，则 $\boldsymbol{A}$ 的列向量**线性相关**

$$\sum_{n=1}^{N}x_n\boldsymbol{a}_n=\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0} \tag{G-2}$$

其中 $x\triangleq[x_{1,\cdots},\ x_N]^{\mathrm{T}}$，$\boldsymbol{0}\in\mathcal{R}^M$ 是全零的列向量。否则，它们线性无关。

类似的定义适用于 $\boldsymbol{A}$ 的行向量。

定义(秩) 矩阵 $\boldsymbol{A}$ 的**秩** r 是线性无关的列的数目，也是线性无关的行的数目。

显然 $r\leqslant\min\{M,N\}$。后面的一系列定义中，假设 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$ 是 $M=N$ 的方阵。

定义(对称矩阵) 如果 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$，$\boldsymbol{A}$ 是对称矩阵；如果 $\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$，$\boldsymbol{A}$ 是反对称矩阵。

对于具有合适行列数的矩阵：

$$(\boldsymbol{A}_1\boldsymbol{A}_2)^{\mathrm{T}}=\boldsymbol{A}_2^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}} \tag{G-3}$$

定义(非奇异) 如果矩阵 $\boldsymbol{A}$ 存在逆矩阵 $\boldsymbol{A}^{-1}$，

$$\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I} \tag{G-4}$$

则 $\boldsymbol{A}$ **非奇异**。其中 $\boldsymbol{I}\in\mathcal{R}^{N\times N}$ 是单位矩阵。如果 $\boldsymbol{A}$ 没有逆矩阵，则**奇异**。

非奇异矩阵 $\boldsymbol{A}$ 具有以下性质：

- $(\boldsymbol{A}^{\mathrm{T}})^{-1}=(\boldsymbol{A}^{-1})^{\mathrm{T}}\triangleq\boldsymbol{A}^{-T}$
- $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ 具有唯一解 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{y}$。
- 行列式非零：$\det(\boldsymbol{A})\neq 0$
- 它具有满秩 $r=N$，并且行和列都是线性无关的。
- 对于合适行列数的非奇异矩阵：

$$(\boldsymbol{A}_1\boldsymbol{A}_2)^{-1}=\boldsymbol{A}_2^{-1}\boldsymbol{A}_1^{-1} \tag{G-5}$$

定义(正定) 如果对于任意非零的 $\boldsymbol{x}\in\mathcal{R}^N$，都有 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}>0$，则矩阵 $\boldsymbol{A}$ 正定(Positive Definite，PSD)。同样地，如果对于任意非零的 $\boldsymbol{x}\in\mathcal{R}^N$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\geqslant 0$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}<0$ 或 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\leqslant 0$ 矩阵 $\boldsymbol{A}$ 分别半正定、负定或半负定。

定义(Toeplitz 和循环) 如果矩阵 $\boldsymbol{A}$ 中任一条平行于主对角线的元素相同，则其是 Toeplitz 矩阵。如果行向量的每个元素都是前一行向量的元素平移一个位置得到的，该 Toeplitz 矩阵是循环矩阵。

一个 Toeplitz 矩阵最多含有 N 个不同元素：任何行或列的元素都能指定整个矩阵。

定义(幂等) 当 $\boldsymbol{A}^2=\boldsymbol{A}$ 时，矩阵 $\boldsymbol{A}$ 是**幂等矩阵**。

定义(正交) 如果 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{I}$，从而 $\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{T}}$，矩阵 $\boldsymbol{A}$ 是**正交矩阵**。

定义(幂零) 如果存在 $K\in\mathcal{N}$，使得 $\boldsymbol{A}^{K}=\boldsymbol{0}$，则矩阵 $\boldsymbol{A}$ 是**幂零矩阵**。

所有主对角线元素是零的三角矩阵都是幂零矩阵。然而，这不是必要条件；存在所有元素非零的幂零矩阵，但他们必须满足一些条件，使得当矩阵幂变为 K 时，所有项最终消去。

例 G-1 随机序列向量 $\boldsymbol{X}[k]$ 的自相关矩阵定义为如下期望：

$$\boldsymbol{R}_{XX}\triangleq\mathcal{E}[\boldsymbol{X}[k]\boldsymbol{X}^{\mathrm{T}}[k]] \tag{G-6}$$

其中 $\boldsymbol{X}[k]\triangleq[X[k],\cdots,X(k-N+1]]^{\mathrm{T}}$ 包含序列的 N 个样本。它是一个对称矩阵：

$$\boldsymbol{R}_{XX}^{\mathrm{T}}=\mathcal{E}^{\mathrm{T}}[\boldsymbol{X}[k]\boldsymbol{X}^{\mathrm{T}}[k]]=\mathcal{E}[\boldsymbol{X}[k]\boldsymbol{X}^{\mathrm{T}}[k]]=\boldsymbol{R}_{XX} \tag{G-7}$$

其特征值是实数。因为对于每一个非零向量 $\boldsymbol{v}$，都有

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{R}_{XX}\boldsymbol{v}=\mathcal{E}[\boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}[k]\boldsymbol{X}^{\mathrm{T}}[k]\boldsymbol{v}]=\mathcal{E}[(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}[k])^{2}]\geqslant 0 \tag{G-8}$$

自相关矩阵也是半正定的。其特征值非负。观察发现 $\boldsymbol{R}_{XX}$ 也是 Toeplitz 矩阵：

$$\begin{aligned}\boldsymbol{R}_{XX}&=\begin{bmatrix}\mathcal{E}[X^{2}[k]] & \mathcal{E}[X[k]X[k-1]] & \cdots & \mathcal{E}[X[k]X[k-N+1]]\\ \mathcal{E}[X[k-1])X[k]] & \mathcal{E}[X^{2}[k-1]] & & \\ \vdots & & \ddots & \vdots\\ \mathcal{E}[X[k-N+1]X[k]] & & \cdots & \mathcal{E}[X^{2}[k-N+1]]\end{bmatrix}\\ &=\begin{bmatrix}R_{XX}[0] & R_{XX}[1] & \cdots & R_{XX}[N-1]\\ R_{XX}[1] & R_{XX}[0] & & \\ \vdots & & \ddots & \vdots\\ R_{XX}[N-1] & & \cdots & R_{XX}[0]\end{bmatrix}\end{aligned} \tag{G-9}$$

其中，$R_{XX}[m]\triangleq\mathcal{E}[X(k]X[k+m]]=R_{XX}[-m]$。这种类型的矩阵出现在第11、12章的最优滤波和自适应滤波的应用中。◀

假设 $\boldsymbol{A}\in\mathcal{C}^{N\times N}$ 是一个复数方阵。

定义(Hermitian) 如果 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，矩阵 $\boldsymbol{A}$ 是 **Hermitian 矩阵**；如果 $\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$，$\boldsymbol{A}$ 是反 Hermitian 矩阵。

上标 ^{H} 表示复共轭转置(也称为 Hermitian 转置)。对于具有合适行列数的矩阵：

$$(\boldsymbol{A}_1\boldsymbol{A}_2)^{\mathrm{H}}=\boldsymbol{A}_1^{\mathrm{H}}\boldsymbol{A}_2^{\mathrm{H}} \tag{G-10}$$

定义(酉) 如果 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{I}$，从而 $\boldsymbol{A}^{-1}=\boldsymbol{A}^{\mathrm{H}}$，矩阵 $\boldsymbol{A}$ 是**酉矩阵**。

复矩阵的 Hermitian 和酉分别对应实矩阵的对称和正交的定义。

G.2 四则子空间

再考虑秩为 r 的实矩阵 $\boldsymbol{A}\in\mathcal{R}^{M\times N}$。

定义(列空间) 矩阵 $\boldsymbol{A}$ 的**列空间**的维数是 r，并且包含列向量的所有线性组合。它也称为 $\boldsymbol{A}$ 的值域。

在下面的表达式中，$\boldsymbol{y}\in\mathcal{R}^{M}$ 属于 $\boldsymbol{A}$ 的列空间：

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}=\sum_{n=1}^{N}x_n\boldsymbol{a}_n \tag{G-11}$$

其中 $\boldsymbol{x}\triangleq\{x_1,\ldots,x_N\}^{\mathrm{T}}$，$\{\boldsymbol{a}_n\}$ 是 $\boldsymbol{A}$ 的列。

定义(零空间) 矩阵 $\boldsymbol{A}$ 的**零空间**的维数是 $N-r$，并且包括所有满足下列条件的向量 $\boldsymbol{x}$：

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0} \tag{G-12}$$

其中 $\boldsymbol{0}\in\mathcal{R}^{M}$。零空间也称为 $\boldsymbol{A}$ 的核，并且其维数称为零化度。

定义(行空间) 矩阵 $\boldsymbol{A}$ 的**行空间**的维数是 r，并且包含行向量的所有线性组合。

在下面的表达式中，$\boldsymbol{y}\in\mathcal{R}^{M}$ 属于 $\boldsymbol{A}$ 的行空间：

$$y = A^{T}x = \sum_{m=1}^{M} x_m \tilde{a}_m \tag{G-13}$$

其中 $x \triangleq [x_1, \cdots, x_M]^{T}$，$\{\tilde{a}_m^{T}\}$是 A 的行。A 的行空间是 A^{T} 的列空间。

定义（左零空间） 矩阵 A 的**左零空间**的维数是 $M-r$，并且包含所有满足下列条件的向量 x：

$$A^{T}x = 0 \tag{G-14}$$

其中 $0 \in \mathcal{R}^M$。A 的左零空间是 A^{T} 的零空间。

G.3 特征分解

在本节中，我们考虑一种特殊类型的矩阵分解——特征分解。它对于研究线性系统的特性非常有用。G.4 节总结了其他类型的矩阵分解。根据特征分解产生的对角矩阵，也称为 A 的对角形式。

定义（特征值和特征向量）方阵 $A \in \mathcal{R}^{N \times N}$ 的**特征值**满足下面的公式：

$$Aq = \lambda q \tag{G-15}$$

非零向量 q 称为**特征向量**。

观察发现式(G-15)可改写为$(A-\lambda I)q=0$，表明 q 属于 $A-\lambda I$ 的零空间。特征值 λ 保证该零空间非空，这意味着 $A-\lambda I$ 是奇异矩阵。由于奇异矩阵的行列式是零，我们可以通过求解下面的特征方程得到特征值：

$$\det(A - \lambda I) = 0 \tag{G-16}$$

直接由式(G-16)得到关于 λ 的 n 次多项式：

$$\lambda^N + \alpha_{N-1}\lambda^{N-1} + \cdots + \alpha_1\lambda + \alpha_0 = 0 \tag{G-17}$$

其中系数$\{\alpha_n\}$由 A 的元素决定。关于 λ 的方程有 N 个解。然而它们有可能是重根，有可能是零，也有可能是复数。在下面的讨论中，假定特征值是实数（如果矩阵是对称矩阵）。

如果 λ_n 是式(G-16)的一个根，对应的特征向量 q_n 通过求解式(G-15)得到：

$$(A - \lambda_n I)q_n = 0 \tag{G-18}$$

由于 λ_n 是特征值，$A-\lambda_n I$ 奇异，它的零空间非空，q_n 是零空间的任意向量。特征向量不唯一：如果 q_n 是一个特征向量，则任意非零 $c \in \mathcal{R}$对应的 cq_n 满足式(G-15)，它也是特征向量。为方便起见，令 q_n 具有单位范数：$q_n^{T}q_n \triangleq ||q_n||^2 = 1$。

定理 G-1 （特征分解）。如果方阵 $A \in \mathcal{R}^{N \times N}$ 是满秩 $r=N$，那么有 N 个线性无关的特征向量，A 可以分解如下：

$$A = Q\Lambda Q^{-1} \tag{G-19}$$

其中对角矩阵 $\Lambda \in \mathcal{R}^{N \times N}$ 包含特征值，矩阵 $Q \in \mathcal{R}^{N \times N}$ 的列是特征向量。如果 A 是一个对称矩阵，则 $Q^{-1} = Q^{T}$。

证明：定义特征向量矩阵

$$Q \triangleq [q_1, \cdots, q_N] \tag{G-20}$$

和特征值对角矩阵

$$\Lambda \triangleq \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & \lambda_N \end{bmatrix} \tag{G-21}$$

根据式(G-15)的定义，逐列排列特征向量：

$$A[q_1, \cdots, q_N] = [\lambda_1 q_1, \cdots, \lambda_N q_N] \tag{G-22}$$

上式可以写成

$$AQ = Q\Lambda \tag{G-23}$$

因为特征向量线性无关，Q^{-1} 存在，(G-19) 得证。如果 A 是对称矩阵，$A^{T} =$

$(Q\Lambda Q^{-1})^{\mathrm{T}}=Q^{-\mathrm{T}}\Lambda Q^{\mathrm{T}}$ 意味着 $Q^{-1}=Q^{\mathrm{T}}$，从而完成了证明。

前面特征向量的定义是最常用的定义之一，虽然严格意义上称它们为*右特征向量*。我们也可以写出特征方程的*左特征向量*，如下所示：

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}=\lambda\boldsymbol{v}^{\mathrm{T}} \tag{G-24}$$

事实证明，特征值与前面结果相同，但一般来说，左和右特征向量不等。将左特征向量逐行排列：

$$\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{v}_N^{\mathrm{T}}\end{bmatrix}\boldsymbol{A}=\begin{bmatrix}\lambda_1\boldsymbol{v}_1^{\mathrm{T}}\\ \vdots\\ \lambda_N\boldsymbol{v}_N^{\mathrm{T}}\end{bmatrix}\Rightarrow\boldsymbol{V}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\Rightarrow\boldsymbol{A}=\boldsymbol{V}^{-\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{G-25}$$

其中 $\boldsymbol{\Lambda}$ 是前面定义的特征值矩阵，$\boldsymbol{V}\triangleq[\boldsymbol{v}_1,\cdots,\boldsymbol{v}_N]$。最后一个结果假设左特征向量是线性无关的。$\boldsymbol{V}^{\mathrm{T}}$ 的行是 $\boldsymbol{A}$ 的左特征向量。如果 $\boldsymbol{A}$ 是一个对称矩阵，则左和右特征向量是彼此的转置，$(\boldsymbol{V}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{Q}\Rightarrow\boldsymbol{V}=\boldsymbol{Q}$，并且

$$\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}=\boldsymbol{A} \tag{G-26}$$

接下来我们将介绍奇异值分解(SVD)，它将特征分解推广到长方矩阵，这里没有给出证明。

定理 G-2 (SVD)。*长方矩阵 $\boldsymbol{A}\in\mathcal{R}^{M\times N}$ 可以进行如下分解：*

$$\boldsymbol{A}=\boldsymbol{U}\sum\boldsymbol{V}^{\mathrm{T}} \tag{G-27}$$

其中 $\boldsymbol{\Sigma}\in\mathcal{R}^{M\times N}$ 是对角线元素非负的对角矩阵，且对角线元素是 $\boldsymbol{A}$ 的奇异值。$\boldsymbol{U}\in\mathcal{R}^{M\times M}$ 是正交矩阵，它的列是左奇异向量，而 $\boldsymbol{V}\in\mathcal{R}^{N\times N}$ 也是正交矩阵，它的列是右奇异向量。

G.4 LU、LDU 和 Cholesky 分解

在介绍正定矩阵的 Cholesky 分解前，先总结其他矩阵分解，这里没有给出证明。

定义(顺序主子式) *矩阵 $\boldsymbol{A}\in\mathcal{R}^{M\times N}$ 的 n 阶顺序主子式 $\boldsymbol{A}_n\in\mathcal{R}^{n\times n}$ 是一个方阵，通过保留 $\boldsymbol{A}$ 的前 n 行和前 n 列得到。*

当然，前面的定义假设 $n\leqslant\min(M,N)$。

定理 G-3 (LU 分解)。*当 $n\leqslant\min(M-1,N)$ 时，如果 $\boldsymbol{A}\in\mathcal{R}^{M\times N}$ 的所有顺序主子式 $\boldsymbol{A}_n\in\mathcal{R}^{M\times N}$ 都非奇异，则*

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{U} \tag{G-28}$$

其中，$\boldsymbol{L}\in\mathcal{R}^{M\times M}$ 是主对角线元素为 1 的下三角矩阵，$\boldsymbol{U}\in\mathcal{R}^{M\times N}$ 是一个上三角矩阵。

需要注意的是，$\boldsymbol{L}$ 永远是方阵，如果 $\boldsymbol{A}$ 不是正阵，则 $\boldsymbol{U}$ 是长方矩阵。主对角线元素为 1 的下(上)三角矩阵称为单位下(上)三角矩阵。

定理 G-4 (LDU 分解)。*如果方阵 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$ 的所有顺序主子式 $\boldsymbol{A}_n$ 都非奇异，则*

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{D}\boldsymbol{U} \tag{G-29}$$

其中 $\boldsymbol{L}\in\mathcal{R}^{N\times N}$ 是一个单位下三角矩阵，$\boldsymbol{U}\in\mathcal{R}^{N\times N}$ 是一个单位上三角矩阵，$\boldsymbol{D}\in\mathcal{R}^{N\times M}$ 是一个对角矩阵。

如果 $\boldsymbol{A}$ 正定，则 $\boldsymbol{D}$ 的对角元素都是正的。

定理 G-5 (Cholesky 分解)。*如果 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$ 是对称正定矩阵，则*

$$\boldsymbol{A}=\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} \tag{G-30}$$

其中 $\boldsymbol{B}\in\mathcal{R}^{N\times N}$ 是对角元素为正的下三角矩阵。

由于在 LDU 分解中，$\boldsymbol{D}$ 的对角元素都为正，显然 $\boldsymbol{B}=\boldsymbol{L}\boldsymbol{D}^{1/2}$，其中 $\boldsymbol{D}^{1/2}$ 是 $\boldsymbol{D}$ 的平方根，即，$\boldsymbol{D}=(\boldsymbol{D}^{1/2})^2$。

例 G-2 考虑对称正定矩阵

$$\boldsymbol{A}=\begin{bmatrix}2 & 1 & 1/2\\ 1 & 2 & 1\\ 1/2 & 1 & 2\end{bmatrix} \tag{G-31}$$

它具有以下 LU 分解：

$$\boldsymbol{L}=\begin{bmatrix}1 & 0 & 0\\ 1/2 & 1 & 0\\ 1/4 & 1/2 & 1\end{bmatrix},\boldsymbol{U}=\begin{bmatrix}2 & 1 & 1/2\\ 0 & 3/2 & 3/4\\ 0 & 0 & 3/2\end{bmatrix} \tag{G-32}$$

其 LDU 分解包括对角矩阵

$$\boldsymbol{D}=\begin{bmatrix}2 & 0 & 0\\ 0 & 3/2 & 0\\ 0 & 0 & 3/2\end{bmatrix} \tag{G-33}$$

其 Cholesky 分解包含下三角矩阵

$$\boldsymbol{B}=\begin{bmatrix}1.4142 & 0 & 0\\ 0.7071 & 1.2247 & 0\\ 0.3536 & 0.6124 & 1.2247\end{bmatrix} \tag{G-34}$$

出于完整性，也给出 $\boldsymbol{A}$ 的特征分解：

$$\boldsymbol{Q}=\begin{bmatrix}-0.4544 & -0.7071 & 0.5418\\ 0.7662 & 0 & 0.6526\\ -0.4544 & 0.7071 & 0.5418\end{bmatrix},\boldsymbol{\Lambda}=\begin{bmatrix}0.8139 & 0 & 0\\ 0 & 1.5 & 0\\ 0 & 0 & 3.6861\end{bmatrix} \tag{G-35}$$ ◀

G.5　雅可比矩阵及雅可比

令 $y_m=g_m(x_1,\cdots,x_N)$ 是从 $\{x_1,\cdots,x_N\}$ 到 $\{y_1,\cdots,y_M\}$ 的 M 个实值函数。

定义(雅可比矩阵)　**雅可比矩阵** $\boldsymbol{J}(x_1,\cdots,x_N)\in\mathcal{R}^{M\times N}$ 包含以下偏微分：

$$\boldsymbol{J}(x_1,\cdots,x_N)\triangleq\begin{bmatrix}\partial y_1/\partial x_1 & \cdots & \partial y_1/\partial x_N\\ \vdots & \ddots & \vdots\\ \partial y_M/\partial x_1 & \cdots & \partial y_M/\partial x_N\end{bmatrix} \tag{G-36}$$

其中 $\boldsymbol{J}(x_1,\cdots,x_N)$ 的参数决定了独立变量，因此矩阵的元素是 $\{x_1,\cdots,x_N\}$ 的函数。

假定 $\boldsymbol{J}(x_1,\cdots,x_N)\in\mathcal{R}^{N\times N}$ 是一个方阵。

定义(雅可比行列式)　方阵 $\boldsymbol{J}(x_1,\cdots,x_N)$ 的雅可比行列式(简称为雅可比)是如下行列式：

$$J(x_1,\cdots,x_N)\triangleq\det(\boldsymbol{J}(x_1,\cdots,x_N))=|\boldsymbol{J}(x_1,\cdots,x_N)| \tag{G-37}$$

当 $M=N=2$ 时，$y_1=g_1(x_1,x_2)$、$y_2=g_2(x_1,x_2)$。可以将同一组变量的两个雅可比定义为：

$$J(x_1,x_2)=\begin{vmatrix}\partial y_1/\partial x_1 & \partial y_1/\partial x_2\\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2\end{vmatrix},J(y_1,y_2)=\begin{vmatrix}\partial x_1/\partial y_1 & \partial x_1/\partial y_2\\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2\end{vmatrix} \tag{G-38}$$

同样的，参数决定了独立变量。可以直接验证以下性质：

$$J(x_1,x_2)J(y_1,y_2)=1 \tag{G-39}$$

当积分中存在变量的变换时，通常利用雅可比 $|J|$ 的绝对值简化积分运算。令 $x_1=h_1(y_1,y_2)$、$x_2=h_2(y_1y_2)$ 是反变换。则函数 $c(x_1,x_2)$ 的积分可以由两种不同方法计算：

$$\begin{aligned}\iint c(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2&=\iint c(h_1(y_1,y_2),h_2(y_1,y_2))|J(y_1,y_2)|\mathrm{d}y_1\mathrm{d}y_2\\&=\iint c(h_1(y_1,y_2),h_2(y_1,y_2))|J(x_1,x_2)|^{-1}\mathrm{d}y_1\mathrm{d}y_2\end{aligned} \tag{G-40}$$

第二个等式利用了式(G-39)。需要注意的是，已经将逆变换代入 $c(x_1,x_2)$，右侧积分限必须由 $\{y_1,y_2\}$ 定义。第二个等式大家可能更熟悉，其雅可比基于变换 $g_1(x_1,x_2)$ 和 $g_2(x_1,x_2)$，通常在问题描述时已经给定了。但是，第一个等式计算比较方便，因为逆变换 $h_1(y_1,y_2)$、$h_2(y_1,y_2)$ 更容易微分。

第 4 章介绍了在计算随机变量变换的联合概率密度函数时，雅可比比较有用。例 G-3

利用雅可比计算一个定积分。

例G-3 应用更广泛的变换之一是直角到极坐标的转换：

$$x = r\cos(\phi),\quad y = r\sin(\phi) \tag{G-41}$$

其

$$J(r,\phi) = \begin{vmatrix} \partial x/\partial r & \partial x/\partial\phi \\ \partial y/\partial r & \partial y/\partial\phi \end{vmatrix} = \begin{vmatrix} \cos(\phi) & -r\sin(\phi) \\ \sin(\phi) & r\cos(\phi) \end{vmatrix}$$
$$= r\cos^2(\phi) + r\sin^2(\phi) = r \tag{G-42}$$

逆变换是

$$x^2 + y^2 = r^2\cos^2(\phi) + r^2\sin^2(\phi) \Rightarrow r = \sqrt{x^2 + y^2} \tag{G-43}$$

$$y/x = r\sin(\phi)/r\cos(\phi) \Rightarrow \phi = \arctan(y/x) \tag{G-44}$$

虽然式(G-42)的雅可比计算简单，但为了完整性，也提供了另一个雅可比：

$$J(x,y) = \begin{vmatrix} \partial r/\partial x & \partial r/\partial y \\ \partial\phi/\partial x & \partial\phi/\partial y \end{vmatrix} = \begin{vmatrix} x/\sqrt{x^2+y^2} & y/\sqrt{x^2+y^2} \\ -y/(x^2+y^2) & x/(x^2+y^2) \end{vmatrix}$$
$$= 1/\sqrt{x^2+y^2} = 1/r = J^{-1}(r,\phi) \tag{G-45}$$

假设要计算下列定积分：

$$\int_0^a \int_0^{\sqrt{a^2-y^2}} (x^2 + y^2)\,\mathrm{d}x\mathrm{d}y \tag{G-46}$$

在上面的变换中使用以下替换：

$$x^2 + y^2 = r^2,\quad \mathrm{d}x\mathrm{d}y = J(r,\phi)\mathrm{d}r\mathrm{d}\phi = r\mathrm{d}r\mathrm{d}\phi \tag{G-47}$$

r 和 ϕ 的下限均为零，并且上限是

$$r = \sqrt{x^2+y^2}\,\big|_{x=\sqrt{\alpha^2-y^2}} = \sqrt{\alpha^2 - y^2 + y^2} = \alpha$$
$$\phi = \arctan(y/\sqrt{\alpha^2 - y^2})\,\big|_{y=\alpha} = \pi/2 \tag{G-48}$$

从而

$$\int_0^a \int_0^{\sqrt{a^2-y^2}} (x^2 + y^2)\,\mathrm{d}x\mathrm{d}y = \int_0^{\pi/2}\int_0^a r^3\,\mathrm{d}r\mathrm{d}\phi = \alpha^4\pi/8 \tag{G-49}$$ ◀

G.6 KRONECKER 和 SCHUR 积

定义 (Kronecker 积)矩阵 $\boldsymbol{A} \in \mathcal{R}^{M\times N}$ 和 $\boldsymbol{B} \in \mathcal{R}^{P\times Q}$ 的 Kronecker 积是一个 $MP \times NQ$ 的分块矩阵：

$$\boldsymbol{A} \otimes \boldsymbol{B} \triangleq \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1N}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{M1}\boldsymbol{B} & \cdots & a_{MN}\boldsymbol{B} \end{bmatrix} \tag{G-50}$$

其中左上分块是

$$a_{11}\boldsymbol{B} = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{11}b_{1Q} \\ \vdots & \ddots & \vdots \\ a_{11}b_{P1} & \cdots & a_{11}b_{PQ} \end{bmatrix} \tag{G-51}$$

其他分块类似。

虽然 Kronecker 积满足结合律，但不满足交换律：$\boldsymbol{A}\otimes\boldsymbol{B} \neq \boldsymbol{B}\otimes\boldsymbol{A}$。还要注意的是 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 可以是向量：如果 $\boldsymbol{A}\rightarrow\boldsymbol{a}\in\mathcal{R}^M$ 是一个列向量，$\boldsymbol{B}\rightarrow\boldsymbol{b}^{\mathrm{T}}\in\mathcal{R}^N$ 是一个行向量，则 $\boldsymbol{A}\otimes\boldsymbol{B}\in\mathcal{R}^{M\times N}$ 是一个矩阵。对于这种情况，Kronecker 积简化为标准向量外积 $\boldsymbol{ab}^{\mathrm{T}}$。

定义(Schur 积) 对于矩阵 $\boldsymbol{A}\in\mathcal{R}^{M\times N}$ 和 $\boldsymbol{B}\in\mathcal{R}^{M\times N}$，通过对应元素分别相乘得到 Schur 积：

$$\boldsymbol{A}\circ\boldsymbol{B}\triangleq\begin{bmatrix}a_{11}b_{11} & \cdots & a_{1N}b_{1N}\\ \vdots & \ddots & \vdots\\ a_{M1}b_{M1} & \cdots & a_{MN}b_{MN}\end{bmatrix} \tag{G-52}$$

因此 $\boldsymbol{A}\circ\boldsymbol{B}\in\mathcal{R}^{M\times N}$。它也称作 Hadamard 积。

Schur 积满足交换律和结合律。

G.7 迹和行列式的性质

定义(迹) 方阵 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$的迹是

$$\mathrm{tr}[\boldsymbol{A}]\triangleq\sum_{n=1}^{N}a_{nn} \tag{G-53}$$

其中$\{a_{nn}\}$是 $\boldsymbol{A}$ 的对角线元素。

迹具有以下性质：

- 特征值$\{\lambda_n\}$：

$$\mathrm{tr}[\boldsymbol{A}]=\sum_{n=1}^{N}\lambda_n,\quad \mathrm{tr}[\boldsymbol{A}^K]=\sum_{n=1}^{N}\lambda_n^K \tag{G-54}$$

- 加法和标量$\{c,d\}$的乘法：

$$\mathrm{tr}[c\boldsymbol{A}+d\boldsymbol{B}]=c\,\mathrm{tr}[\boldsymbol{A}]+d\,\mathrm{tr}[\boldsymbol{B}] \tag{G-55}$$

- 交换率：$\mathrm{tr}[\boldsymbol{AB}]=\mathrm{tr}[\boldsymbol{BA}]$以及

$$\mathrm{tr}[\boldsymbol{ABC}]=\mathrm{tr}[\boldsymbol{BCA}]=\mathrm{tr}[\boldsymbol{CAB}] \tag{G-56}$$

$$\neq\mathrm{tr}[\boldsymbol{CBA}]\text{ 或 }\mathrm{tr}[\boldsymbol{ACB}]\text{ 或 }\mathrm{tr}[\boldsymbol{BAC}] \tag{G-57}$$

- Kronecker 积：

$$\mathrm{tr}[\boldsymbol{A}\otimes\boldsymbol{B}]=\mathrm{tr}[\boldsymbol{A}]\mathrm{tr}[\boldsymbol{B}] \tag{G-58}$$

- 相似不变：对于可逆矩阵 $\boldsymbol{P}$

$$\mathrm{tr}[\boldsymbol{P}^{-1}\boldsymbol{AP}]=\mathrm{tr}[\boldsymbol{A}] \tag{G-59}$$

- 半正定矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的限：

$$0\leqslant\mathrm{tr}[(\boldsymbol{AB})^K]\leqslant(\mathrm{tr}[\boldsymbol{A}]\mathrm{tr}[\boldsymbol{B}])^K \tag{G-60}$$

方阵 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$的行列式有以下性质：

- 转置：$\det(\boldsymbol{A})=\det(\boldsymbol{A}^{\mathrm{T}})$。
- 逆：$\det(\boldsymbol{A}^{-1})=1/\det(\boldsymbol{A})$。
- 乘以一个常数 c：$\det(c\boldsymbol{A})=c^N\det(\boldsymbol{A})$。
- 分配：如果 $\boldsymbol{B}\in\mathcal{R}^{N\times N}$，$\det(\boldsymbol{AB})=\det(\boldsymbol{A})\det(\boldsymbol{B})$。

G.8 矩阵求逆引理

矩阵的下列性质，也称为 Sherman-Morrison-Woodbury 公式，在第 11 章推导递归最小二乘(Recursive Least-Squares，RLS)算法时是非常有用的。

引理 G-1 (矩阵求逆)。对于非奇异方阵 $\boldsymbol{B}$ 和 $\boldsymbol{D}$：

$$(\boldsymbol{B}+\boldsymbol{UDV}^{\mathrm{T}})^{-1}=\boldsymbol{B}^{-1}-\boldsymbol{B}^{-1}\boldsymbol{U}(\boldsymbol{D}^{-1}+\boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}^{-1} \tag{G-61}$$

其中，矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 不一定是方阵。

证明参见习题 11-17。

当 $\boldsymbol{U}\to\boldsymbol{u}$ 和 $\boldsymbol{V}\to\boldsymbol{v}$ 是列向量时，矩阵求逆引理简化为 Sherman-Morrison 公式：

$$(\boldsymbol{B}+d\boldsymbol{uv}^{\mathrm{T}})^{-1}=\boldsymbol{B}^{-1}-\boldsymbol{B}^{-1}\boldsymbol{uv}^{\mathrm{T}}\boldsymbol{B}^{-1}/(d^{-1}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{u}) \tag{G-62}$$

由于 $\boldsymbol{D}\to d$ 必然是一个标量，除数变成标量 $d^{-1}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{u}$，这时计算相当简单。这是用来导出 RLS 算法的套路。

G.9 柯西-施瓦兹不等式

同样维数的向量对应的柯西-施瓦兹不等式是

$$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y})^2 \leqslant \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} = \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \tag{G-63}$$

当且仅当 $\boldsymbol{x}=c\boldsymbol{y}$，其中 $c\in\mathcal{R}$ 非零时等号成立。等效的不等式是

$$|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leqslant \|\boldsymbol{x}\|\|\boldsymbol{y}\| \tag{G-64}$$

随机变量期望的柯西-施瓦兹不等式见附录F.

G.10 微分

设 $\boldsymbol{A}$ 是任意非零矩阵，$\boldsymbol{x}$ 和 $\boldsymbol{v}$ 是具有适当维数的列向量。当对矩阵或向量计算式关于列向量求导时，以下结果非常有用：

$$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x} = \boldsymbol{I}, \quad \frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x} = \frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{v}, \quad \frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 2\boldsymbol{x} \tag{G-65}$$

$$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x} = 2\boldsymbol{A}\boldsymbol{x}(\boldsymbol{A}\text{ 为对称矩阵}) \tag{G-66}$$

$$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v} = \boldsymbol{A}\boldsymbol{v}, \quad \frac{\partial}{\partial \boldsymbol{v}}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \tag{G-67}$$

$$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}, \quad \frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{A} \tag{G-68}$$

例 G-4 考虑以下线性方程：

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{G-69}$$

如果方阵 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$ 是满秩 $r=N$，那么它有逆矩阵，解是 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{y}$。然而，如果 $\boldsymbol{A}\in\mathcal{R}^{M\times N}$ 为长方形，$M>N$，方程组超定(如果 $M\gg N$，$\boldsymbol{A}$ 是一个高窄矩阵)。对于这种情况，我们力求减少以下LS代价函数：

$$\xi_{\mathrm{LS}} \triangleq \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2 = (\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} \tag{G-70}$$

对 ξ_{LS} 关于 $\boldsymbol{x}$ 进行微分并令结果为0，得到

$$2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_{\mathrm{LS}} - 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0} \tag{G-71}$$

LS解是

$$\boldsymbol{x}_{\mathrm{LS}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{y} \tag{G-72}$$

其中，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\in\mathcal{R}^{N\times N}$ 是方阵且满秩。第9章介绍了LS技术。矩阵

$$\boldsymbol{A}^{+} \triangleq (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A} \tag{G-73}$$

是Moore-Penrose广义逆。 ◀

例 G-5 考虑下面的约束优化问题：

$$\boldsymbol{x}_{\mathrm{o}} = \arg\max_{\boldsymbol{x}}[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \lambda(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{c}-c)] \tag{G-74}$$

其中 λ 是拉格朗日乘数(不要与特征值混淆)，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{c}=c$ 是 $\boldsymbol{x}$ 上的线性约束，c 为标量。令 $\boldsymbol{A}\in\mathcal{R}^{N\times N}$，则 $\boldsymbol{x}$，$\boldsymbol{c}\in\mathcal{R}^{N}$。对该表达式关于 $\boldsymbol{x}$ 和 λ 求导，得到：

$$2\boldsymbol{A}\boldsymbol{x}_0 + \lambda_0\boldsymbol{c} = \boldsymbol{0}, \boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{c} - c = 0 \tag{G-75}$$

(注意，因为结果刚好是约束条件，通常不必关于 λ 求导。) ◀

重新排列第一个等式如下：

$$\boldsymbol{x}_{\mathrm{o}} = -(\lambda_0/2)\boldsymbol{A}^{-1}\boldsymbol{c} \tag{G-76}$$

该表达式左乘 $\boldsymbol{c}^{\mathrm{T}}$ 并利用约束条件，得到

$$c = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}_0 = -(\lambda_0/2)\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{c} \Rightarrow \lambda_0 = -2c/\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{c} \tag{G-77}$$

将该标量代入式(G-75)的第一个等式，得到最终的结果：

$$\boldsymbol{x}_0 = \frac{c\boldsymbol{A}^{-1}\boldsymbol{c}}{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{c}} \tag{G-78}$$

约束条件里的标量 c 控制 $\boldsymbol{x}_0$ 的范数：

$$\| \boldsymbol{x}_0 \|^2 = \frac{c^2 \boldsymbol{c}^{\mathrm{T}} \boldsymbol{A}^{-1} \boldsymbol{c}}{(\boldsymbol{c}^{\mathrm{T}} \boldsymbol{A}^{-1} \boldsymbol{c})^2} = \frac{c^2}{\boldsymbol{c}^{\mathrm{T}} \boldsymbol{A}^{-1} \boldsymbol{c}} \tag{G-79}$$

G.11 复数微分

对于复数矩阵 $\boldsymbol{A} \in C^{M \times N}$ 和适当维数的复数列向量 $\boldsymbol{x}$ 和 $\boldsymbol{v}$：

$$\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x} = \boldsymbol{I}, \quad \frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{x} = \boldsymbol{0} \tag{G-80}$$

其中上标 * 表示复共轭。第二个表达式叫做共轭导数。本书中使用的导数涉及复数，如第13章在波束合成中采用的梯度基于共轭导数。下式给出了复数 $\boldsymbol{A}$、$\boldsymbol{x}$、$\boldsymbol{v}$ 的一些导数：

$$\frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{v}^{\mathrm{H}} \boldsymbol{x} = \boldsymbol{0}, \frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{v} = \boldsymbol{v}, \quad \frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{x} = \boldsymbol{x}, \tag{G-81}$$

$$\frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x} = \boldsymbol{A} \boldsymbol{x}, \quad \frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{v} = \boldsymbol{A} \boldsymbol{v}, \frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{v}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x} = \boldsymbol{0} \tag{G-82}$$

$$\frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{A} \boldsymbol{x} = \boldsymbol{0}, \quad \frac{\partial}{\partial \boldsymbol{x}^*} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} = \boldsymbol{A} \tag{G-83}$$

例 G-6 再利用复数考虑例 G-5 中的优化问题：

$$\begin{aligned} \boldsymbol{x}_0 &= \arg \max_{\boldsymbol{x}} [\boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x} + \lambda^* (\boldsymbol{x}^{\mathrm{H}} \boldsymbol{c} - c) + \lambda (\boldsymbol{c}^{\mathrm{H}} \boldsymbol{x} - c^*)] \\ &= \arg \max_{\boldsymbol{x}} [\boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x} + 2\mathrm{Re}(\lambda^* (\boldsymbol{x}^{\mathrm{H}} \boldsymbol{c} - c))] \end{aligned} \tag{G-84}$$

其中 λ 是一个复值拉格朗日乘数。对这个表达式关于 $\boldsymbol{x}^*$ 和 λ^* 微分，并利用前面给出的准则，得出

$$\boldsymbol{A} \boldsymbol{x}_0 + \lambda_0^* \boldsymbol{c} = 0, \quad \boldsymbol{c}^{\mathrm{H}} \boldsymbol{x}_0 - c = 0 \tag{G-85}$$

推导的其余步骤与例 G-6 相同：

$$\boldsymbol{x}_0 = -\lambda_n^* \boldsymbol{A}^{-1} \boldsymbol{c} \Rightarrow \boldsymbol{c}^{\mathrm{H}} \boldsymbol{x}_0 = -\lambda^* \boldsymbol{c}^{\mathrm{H}} \boldsymbol{A}^{-1} \boldsymbol{c} \tag{G-86}$$

且

$$\lambda_0^* = -c^* / \boldsymbol{c}^{\mathrm{H}} \boldsymbol{A}^{-1} \boldsymbol{c} \Rightarrow \boldsymbol{x}_0 = \frac{c^* \boldsymbol{A}^{-1} \boldsymbol{c}}{\boldsymbol{c}^{\mathrm{H}} \boldsymbol{A}^{-1} \boldsymbol{c}} \tag{G-87}$$

它与实数 $\boldsymbol{x}_0$ 形式类似。

进一步阅读

本附录中的内容可在下列文献中进一步研究：Golub 和 Van Loan(1983)，Haykin(1991)，Kreyszig(1979)，Laub(2005)，Leon(1990)，Schilling 和 Lee(1988)，Strang(1976)。

术语表

符号总结

X	大写字母：随机变量
x	小写字母：随机变量 X 的输出
$\lvert X \rvert$	随机变量 X 的绝对值
$X_{(n)}$	n 阶统计量
$\hat{X}$	X 的估计
$\tilde{X}$	误差随机变量 $X-\hat{X}$
$\boldsymbol{X}$	加粗的大写字母：随机向量
$\boldsymbol{x}$	加粗的小写字母：随机向量 $\boldsymbol{X}$ 的输出
$X[k]$	k 的大写的函数：随机序列
$x[k]$	k 的小写的函数：$X[k]$的现实
$X(t)$	t 的大写的函数：随机过程
$x(t)$	t 的小写的函数：$X[t]$的现实
$X'(t)$	随机过程的导数
$\boldsymbol{X}[k]$	k 的加粗大写函数：随机向量序列
$\boldsymbol{x}[k]$	k 的加粗小写函数：$\boldsymbol{X}[k]$的现实
$\boldsymbol{X}[t]$	t 的加粗大写函数：随机向量过程
$\boldsymbol{x}[t]$	t 的加粗小写函数：$\boldsymbol{X}[t]$的现实
$X(t)*Y(t)$，$X[k]*Y[k]$	（连续，离散）卷积
$X(t)\star Y(t)$，$X[k]\star Y[k]$	（连续，离散）确定的互相关
$X(s)$	拉普拉斯变换
$X(z)$	z 变换
$X(\omega)$	傅里叶变换或随机变量映射
$X(\mathrm{j}\omega)$	离散时间傅里叶变换
$\boldsymbol{X}(\omega)$	随机向量映射
$X^{-1}(B)$，$\boldsymbol{X}^{-1}(B)$	事件 B 的逆象
$\boldsymbol{a}$	列向量
$\lVert\boldsymbol{a}\rVert^2$	平方范数 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$
$\boldsymbol{a}^*$	复共轭
$\boldsymbol{a}^{\mathrm{T}}$	行向量
$\boldsymbol{A}$	非随机矩阵
$\boldsymbol{A}^{\mathrm{T}}$	矩阵的转置
$\boldsymbol{A}^{\mathrm{H}}$	矩阵的共轭转置
$\boldsymbol{A}^{-1}$	矩阵的逆
$\boldsymbol{A}^{-\mathrm{T}}$	矩阵逆的转置
$\boldsymbol{A}^{-\mathrm{H}}$	矩阵逆的共轭转置
$\boldsymbol{A}^{+}$	穆尔-彭罗斯伪逆
$\boldsymbol{A}^{\circ}\boldsymbol{B}$	舒尔积
$\boldsymbol{A}\otimes\boldsymbol{B}$	克罗内克乘积
$\lvert\boldsymbol{A}\rvert$	方阵 $\boldsymbol{A}$ 的行列式
$A\cup B$，$A+B$	集合的并

$A\cap B$, AB	集合的交
$A-B$, $A\setminus B$	集合的差
$A\oplus B$	集合的异或
A^c, $\overline{A}$	集合的补
$A\times B$	集合 A 和 B 的笛卡儿积
$\lvert A\rvert$	集合 A 的基数
$[a, b]$, $[a, b)$, $(a, b]$, (a, b)	实数$\mathcal{R}$的间隔子集
$\{a, \cdots, b\}$	整数 Z 的有限子集
$g'(x)$	一般导数
$g^{(n)}(x)$	n 阶一般导数
$g_E(x)$	偶函数
$g_O(x)$	奇函数
$\binom{n}{x}$	二项式系数
$\binom{n}{x_1, x_2, \cdots, x_m}$	多项式系数
$x!$	阶乘
$x^{(n)}$	升阶乘
$(x)_n$	降阶乘

一般符号

$\mathbf{0}$	0 的列向量或矩阵
$\mathbf{1}$	1 的列向量
$\beth_n$	n 阶贝斯数(无穷集的势)
∞, $-\infty$	无穷大符号(非实数)
$\arg(s)$	复数 s 的复角
$\arg\min_\theta g(x;\theta)$	$g(x;\theta)$ 的 θ 最小化
$\arg\max_\theta g(x;\theta)$	$g(x;\theta)$ 的 θ 最大化
B_n	伯努利数
$B(a,b)$	贝塔函数
$B_y(a,b)$	不完全贝塔函数
B_{3dB}	3dB 带宽
B_{NE}	噪声等效带宽
B_{RMS}	均方根带宽
$B_m[k]$	m 阶反向预测误差
$B(\theta)$	参数 θ 的估计器的偏差或一个天线阵的波束形状
$\mathrm{ceil}(x)$	上取整函数
$\mathrm{cov}[X,Y]$	X 和 Y 的协方差
$\mathrm{cov}[\boldsymbol{X}]$	$\boldsymbol{X}$ 的自协方差矩阵
c	一般位置参数
$c(t)$	累积量母函数
c_X	变异系数
C	康托尔集，每个符号的通道容量，或轮廓积分
$\{C_m\}$	康托尔集的集合序列
C_r	通道容量比率
$C(A)$	集合 A 的计数测度
$C(t,\theta)$	损失函数
C_{XX}, C_{XY}	自协方差，互协方差
$C_{XX}(t_1,t_2)$, $C_{XX}[m,n]$	(连续，离散)自协方差函数
$C_{XY}(t_1,t_2)$, $C_{XY}[m,n]$	(连续，离散)互协方差函数
$\mathbf{C}_{XX}$, $\mathbf{C}_{XY}$	自协方差矩阵，互协方差矩阵

$\widetilde{\boldsymbol{C}}_{XX}$	主对角线全1的 $\boldsymbol{C}_{XX}$
$\det(\boldsymbol{A})$	方阵 $\boldsymbol{A}$ 的行列式
d	均匀线性阵列天线间的间距
$D(x)$	狄里克雷函数
$D(p_X \parallel q_X)$	库尔贝克-莱布勒散度
$\boldsymbol{D}$	对角方阵
e	自然对数的底：2.7182818284590452353 6…
$\mathrm{erf}(x)$	误差函数
$\mathrm{erfc}(x)$	余误差函数
$\mathrm{erfi}(x)$	虚误差函数
$\exp(x)$	指数函数(代替 e^x)
$\mathbf{e}_m$	m 阶单位向量
E	能量或一个集合
E_b	每位的能量
$E[k]$	误差信号
$\mathrm{floor}(x)$	下取整函数
f	单位为 Hz 的频率
f_s	单位为 Hz 的采样频率
$f_X(x)$	概率密度函数(pdf)
$f_{X,Y}(x,y)$	联合 pdf
$f_{X\|Y}(x\|y)$	条件 pdf
F	一个集合
$F_X(x)$	累积分布函数(cdf)
$F^{-1}(x)$	逆 cdf
$F_{X,Y}(x,y)$	联合 cdf
$F_{X\|Y}(x\|y)$	条件 cdf
$F_C(x)$	康托尔函数
$\{F_{C_m}(x)\}$	康托尔函数的函数序列
$F_m[k]$	m 阶前向预测误差
${}_1F_1(a;b;x)$	第一类合流超几何函数
$\{g_n(t)\}$	Gram-Schmidt 基函数
$G_X(t)$	概率母函数
$\boldsymbol{G}[k]$	最小二乘增益向量或卡尔曼滤波器的增益矩阵
H_0, H_n	原假设，第 n 个假设
$h(X)$	连续变量 X 的微分熵
$H(X)$	离散变量 X 的熵
$H(Y\|X)$	条件熵
$H(\omega), H(f), H(s)$	(连续时间)传递函数
$H(\mathrm{j}\omega), H(z)$	(离散时间)传递函数
$H_n(v)$	调和数
$\{H, T\}$	投掷硬币的正面，反面
$H_T(f), H_R(f)$	发送，接收滤波器
$\boldsymbol{H}[k]$	Hessian 矩阵
$\inf(a,b)$	下确界
$i(t)$	电路电流
$I(x_n)$	x_n 的自信息
$I(x_n;y_n)$	x_n 和 y_n 的互信息
$I(X;Y)$	平均互信息
$I(X;Y\|Z)$	条件互信息
$I_{[a,b]}(x)$	连续变量 x 的指示函数

$I_{\{a,b\}}[x]$	离散变量 x 的指示函数
$I_\alpha(x)$	第一类改进的贝塞尔函数
$I_y(a,b)$	正则不完全 β 函数
I_s	电流源
$I_m(s)$	s 的虚部
$\boldsymbol{I}$	单位矩阵
j	$\sqrt{-1}$
$J_\alpha(x)$	第一类贝塞尔函数
J	雅可比
$\boldsymbol{J}$	雅可比矩阵
k	离散时间
$K_\alpha(x)$	第二类贝塞尔函数
$l(\boldsymbol{x})$	似然函数
$L(\boldsymbol{x})$	对数似然函数
$\ln(x)$	自然对数函数
$\log_a(x)$	以 a 为底的对数函数
$L_v(x)$	拉盖尔多项式
L^m	L^m 范数
$L(A)$	集合 A 的勒贝格测度
$L[k]$	随机序列的似然比
$LB(N)$	勒贝格和
$\boldsymbol{L}$	下三角或下移矩阵
$\min\{x,y\}$	最小值函数
$\max\{x,y\}$	最大值函数
m,n	F 分布的参数
m_e	中值
m_o	众数
$m_X(t)$	矩母函数
$M_X(t)$	中心矩母函数
N	变量的数目或自由度
$N_0/2$	白噪声的功率谱密度
$o(x)$	小 o 符号
$O(x)$	大 O 符号
p,q	“成功”概率 p 和“失败”概率 $q \triangleq 1-p$
$p_X[x]$	概率质量函数(pmf)
$p_{X,Y}[x,y]$	联合 pmf
$p_{X\|Y}[x\|y]$	条件 pmf
P	$x(t)$的功率或语音的音高
$P(E)$	事件 E 的概率
$P(E\|F)$	条件概率
$P_X(x)$	$P(X=x)$(不是 pmf)
P_{XX}, P_{XY}	$X(t)$的功率，$X(t)$和 $Y(t)$的互功率
P_e	平均误差概率
$P_{ac}(X\leqslant x)$	绝对连续 cdf
$P_s(X\leqslant x)$	奇异 cdf
$P(\theta)$	到达角估计函数
$\tilde{p}_n^{(k)}$	马尔可夫链返回状态 n 的概率
p_{mn}	马尔可夫链的条件概率
$\boldsymbol{p}$	马尔可夫链的状态概率
$\boldsymbol{P}$	元素为$\{p_{mn}\}$的一步马尔可夫链的状态转移矩阵

$\boldsymbol{P}^{(r)}$	元素为$\{p_{mn}^{(r)}\}$的 r 步马尔可夫链的状态转移矩阵
q_m	第 m 个分位数
$q_X[x]$	用于库尔贝克-莱布勒发散量的 pmf
$\boldsymbol{q}$	特征向量
$\boldsymbol{Q}$	特征向量矩阵
$Q(x)$	Q 函数
$Q_m(a,b)$	马坎 Q 函数
r	一般形状参数，自由度，半径，或复数的幅度
r_0,r_1,r_2	盲均衡算法的增益参数
$\{R,L,C\}$	电阻 R，电感 L，电容 C
R_n	最大似然信号检测的第 n 个判决域
R	黎曼积分或风险函数
$R(N),R_L(N),R_U(N)$	黎曼和、下黎曼和及上黎曼和
RS	黎曼-斯蒂尔切斯积分
$RS(N)$	黎曼-斯蒂尔切斯和
$R(x)$	斜坡函数
R_{XX},R_{XY}	自相关，互相关
$R_{XX}(t_1,t_2),R_{XX}[m,n]$	(连续，离散)自相关函数
$R_{XY}(t_1,t_2),R_{XY}[m,n]$	(连续，离散)互相关函数
$\overline{R}_{XX}$	随机过程 $X(t)X(t+\tau)$的时间平均
$\overline{\overline{R}}_{XX}$	现实 $x(t)x(t+\tau)$的时间平均
$\mathrm{rank}[\boldsymbol{A}]$	矩阵 $\boldsymbol{A}$ 的秩
$\mathrm{rect}(x),\mathrm{rect}[x]$	(连续，离散)矩形函数
$\mathrm{Re}(s)$	s 的实部
$\mathrm{Res}[X(z),p]$	对 $X(z)$估计为 p 的残差
s	复 s 域变量：$s=\sigma+\mathrm{j}\omega$
$\mathrm{sgn}(x)$	正负(或符号)函数
$\mathrm{sinc}(x)$	辛格函数
$\mathrm{stdev}[X]$	X 的标准差
$\sup(a,b)$	上确界
$Sa(x)$	Sa 函数
$S_{XX}(s),S_{XX}(z)$	(s 域，z 域)功率谱密度
$S_{XX}^{+}(z),S_{XX}^{-}(z)$	谱因式分解：极点/零点在单位圆内，在单位圆外
$S_{XX}^{+}(s),S_{XX}^{-}(s)$	谱因式分解：极点/零点在 s 平面的左半区，右半区
$S_{XX}(\omega),S_{XX}(f),S_{XX}(\mathrm{j}\omega)$	(频率域)功率谱密度
$S_{XX}^{D}(s),S_{XX}^{D}(z)$	(s 域，z 域)功率谱分布
$S_{XX}^{D}(\omega),S_{XX}^{D}(f),S_{XX}^{D}(\mathrm{j}\omega)$	(频率域)功率谱分布
S^2	样本方差
t	矩和累积量母函数的连续时间或变量
$\mathrm{tr}[\boldsymbol{A}]$	矩阵 $\boldsymbol{A}$ 的迹
$\mathrm{tri}(x)$	三角函数
T	参数的统计值/估计值或时间长度
$\boldsymbol{T}$	参数向量的统计值/估计值
T_d	脉冲持续时间
T_o	脉冲时间间隔
T_s	采样周期
$\mathrm{var}[X]$	变量 X 的方差
$u(t)$	单位阶跃(赫维赛德)函数
$u[k]$	离散单位阶跃(赫维赛德)函数
$\boldsymbol{U}$	上三角或上移位矩阵
$\mathrm{var}[X]$	变量 X 的方差

V	维塔利集
W	单边带宽
$\boldsymbol{W}$	权重矩阵
$X_{DM}[k]$	$X[k]$的杜布-迈耶可预测分量
$X_{MS}[k]$	$X[k]$的均方可预测分量
$\widetilde{X}(t)$	基带信号
$X_I(t), X_Q(t)$	同相与正交基带分量
$X_q(t)$	量化信号
$\overline{X}$	采样平均值
$\overline{X^n}$	第n时刻采样值
z	复z域变量：$z=r\exp(j\omega)$
z^{-1}	延迟算子：$z^{-1}x[k]=x[k-1]$

希腊符号

α	pdf 尺度参数或算法的步骤参数
α_c	$1-\alpha_c=$置信度(置信区间)
α_s	(假设检验中的)显著性水平
$\alpha_{n,m}$	恒模阵列增益系数
β	贝塔分布的位置、缩放、形状参数，最小均方算法的"泄露"因子，或升余弦脉冲的滚降系数
$\beta_{n,m}$	恒模阵列移位因子
γ	Euler-Mascheroni 常量：0.5772156649015328606O…
γ_1	偏态
γ_2	峰度(过载)
$\gamma(a,b)$	下不完全伽马函数
Γ_m	m阶反射系数
$\Gamma(x)$	伽马函数
$\Gamma(a,b)$	上不完全伽马函数
δ	正值小量
$\delta(x),\delta(x-y),\delta(x,y)$	狄拉克 delta 函数
$\delta[m],\delta[m-n]$	克罗内克 delta 函数
$\delta_\omega(A)$	集合A的狄拉克测度
Δ	矩阵行列式或量化间隔
Δ_{cd}	采样值中一致对的数量与非一致对的数量之差
Δt	时间间隔t/k
Δ_N	时间间隔$(t-t_0)/N$
$\Delta X[k]$	差分序列$X[k]-X[k-1]$
∇	斜率
$\nabla[k]$	斜率向量
ε	非常小的正值
ζ	随机过程采样空间$\mathcal{S}$中的元素
$\zeta(s)$	黎曼 zeta 函数
η	最小均方算法的"泄露"因子，似然比检验的门限，或归一化最小均方算法的参数
θ	参数或角
$\boldsymbol{\theta}$	参数向量
κ_n	n阶累积量
λ	特征值，波长，拉格朗日乘子，pdf 位置/缩放参数，或遗忘因子
$\Lambda(x)$	三角函数，似然比检验
$\boldsymbol{\lambda}$	拉格朗日乘子向量
$\boldsymbol{\Lambda}$	平方对角矩阵

$\mu(A)$	集合 A 的测度
μ	μ 律量化器的均值或参数
μ_n	n 阶原点矩
$\mu_{c,n}$	n 阶中心矩
$\widetilde{\mu}_{c,n}$	n 阶标准化中心矩
$\overline{\mu}_X$	随机过程 $X(t)$的时间平均
$\overline{\overline{\mu}}_X$	现实 $x(t)$的时间平均
ξ	代价函数
π	3.14159265358979323846…
$\{\pi(1),\cdots,\pi(N)\}$	$\{1,\cdots,N\}$ 的置换
$\prod_n$	对下标 n 求乘积
ρ	相关系数
$\overline{\rho}$	采样相关系数
σ	标准差或复数 s 的实部
σ^2	方差
$\sigma(\mathcal{G})$	集合$\mathcal{G}$的 sigma 域
Σ	对角矩阵
$\sum_n$	对下标 n 求和
τ	时间间隔 t_2-t_1 或时间常量
τ_k	肯德尔等级相关系数
τ_S	斯皮尔曼等级相关系数
ϕ	复数量的相位或空集
$\{\phi_n(t)\}$	卡夫南-勒夫扩展基函数
$\Phi_X(\omega)$	(连续 pdf 的)特征函数
$\Phi_X(j\omega)$	(离散 pmf 的)特征函数
χ_N	(N 级自由度)chi 随机变量
χ_N^2	(N 级自由度)卡方随机变量
$\Psi(a;bx)$	第二类合流超几何函数
ω	Ω 的元素或角频率(角度/秒)
ω_s	采样角频率(角度/秒)
Ω	随机变量的采样空间

花体符号

$\mathcal{A}$	域元素的集合
$\mathcal{B}(\mathcal{R})$	博雷尔 σ 域
$\mathcal{C}$	复数域
$\mathcal{E}[X]$	X 的期望值
$\mathcal{E}[X\|Y]$	条件期望值
$\mathcal{E}_Y[\cdot]$	基于 Y 的 pdf 的期望值
$\mathcal{F}$	域或 σ 域
$\mathcal{G}$	σ 域的生成元
$\mathcal{I}(\theta)$	费歇尔信息
$\mathcal{I}(\theta)$	费歇尔信息矩阵
$\mathcal{M}$	最小均方算法的失调
$\mathcal{N}$	自然数集{1, 2, …}
$\mathcal{P}(E)$	集合 E 的幂集
$\mathcal{Q}$	有理数域
$\mathcal{R}$	实数域$(-\infty,\infty)$
$\mathcal{R}^+$	非负实数集$[0,\infty)$

$\mathcal{R}^{-}$ 非正实数集$(-\infty, 0]$
$\mathcal{R}^{N}$ $N\times 1$实数向量
$\mathcal{R}^{M\times N}$ $M\times N$实数矩阵
$\mathcal{S}$ 随机过程的采样空间
$\mathcal{T}$ 时间点集合
$\mathcal{Z}$ 整数集$\{\cdots, -2, -1, 0, 1, 2, \cdots\}$
$\mathcal{Z}^{+}$ 非负整数集$\{0, 1, 2, \cdots\}$
$\mathcal{Z}^{-}$ 非正整数集$\{\cdots, -1, 0\}$

数学符号

$\rightarrow$ 频域变换为或替换为
$\xrightarrow{a.s.}$ 几乎必然收敛于
$\xrightarrow{d}$ 依分布收敛于
$\xrightarrow{L^m}$ 依均值收敛于
$\xrightarrow{m.s.}$ 依均方收敛于
$\xrightarrow{p}$ 依概率收敛于
$\xrightarrow{s}$ 必然收敛于
$\sim$ 按…分布或与…相似
$\Rightarrow$ 可导出
$\triangleq$ 定义为
$\equiv$ 等价于
$\gg$ 远大于
$\ll$ 远小于
$\subset$ 真包含子集
$\subseteq$ 子集
$\in$ 元素属于
$\notin$ 元素不属于

缩略语

a. s. 几乎必然地
cdf 累积分布函数
icdf cdf 的逆
iid 独立同分布
isi 独立平稳增量
pdf 概率密度函数
plim 依概率收敛
pmf 概率质量函数
s. 必然地
s. t. 满足
w. p. 1 以概率 1
A 安培
AM 振幅调制
AOA 到达角
AR 自回归
ARMA 自回归滑动平均
ASK 振幅相移键控
AWGN 加性高斯白噪声
BER 误比特率

BLU	最佳线性无偏
BPSK	二进制相移键控
BSC	二元对称信道
CF	特征函数
CGF	累积量母函数
CLT	中心极限定理
CM	恒模
CMA	恒模算法
CMGF	中心矩母函数
CRLB	克拉美罗下限
dB	分贝
DD	直接判决
DE	微分方程或差分方程
DF	测向
DFT	离散傅里叶变换
DM	杜布-梅耶
DTFT	离散时间傅里叶变换
EM	期望最大化
F	华氏
FFT	快速傅里叶变换
FIR	有限冲激响应
FM	频率调制
FMGF	阶乘矩母函数
FR	强迫响应
FSK	频移键控
FSM	有限状态机
GS	格莱姆-施密特
Hz	赫兹
IIR	无限冲激响应
ISI	码间干扰
KL	库尔贝克-莱布勒
KLE	卡夫南-勒夫扩展
LD	莱文森-杜宾
LMF	四阶最小二乘均值
LMS	最小均方
LoS	视线
LPC	线性预测码
LPF	低通滤波器
LRT	似然比检验
LS	最小二乘法
LTI	线性时不变
MA	滑动平均
MAE	平均绝对误差
MAP	最大后验
MF	匹配滤波器
MGF	矩生成函数
MISO	多输入单输出
ML	最大似然
MMSE	最小均方误差
MoM	矩量法
MQ	四次平均

MQE	四次平均误差
MS	均方
MSE	均方误差
MVDR	最小方差无失真响应
MUSIC	多信号分类
MVUE	最小方差无偏估计
NE	等效杂波
NLMS	归一化最小均方算法
NM	牛顿法
NR	自然响应
NS	噪声子空间
PAM	脉冲幅度调制
PCA	主成分分析
PCM	脉码调制
PCR	主成分回归
PD	正定
PFE	部分分式展开
PGF	概率母函数
PLL	锁相环
PRN	伪随机噪声或伪随机数
PSD	功率谱密度或半正定
PSK	相移键控
QAM	正交幅度调制
QPSK	正交相移键控
RB	罗-布莱克威尔
RC	升余弦
RLC	电路：电阻 R，电感 L，电容 C
RLS	递归最小二乘法
RMS	均方根
ROC	收敛域
SD	最速下降
SINR	信号与干扰加噪声之比
SISO	单输入单输出
SNR	信噪比
SOI	感兴趣的信号
SQNR	信号量化噪声比
SRRC	平方根升余弦
TDL	抽头延迟线
ULA	均匀线阵
UMVU	一致最小方差无偏
V	伏特
V/UV	清音/浊音
WLS	加权最小二乘法
ZF	迫零

参考文献

K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, August 1997.

S. T. Alexander, *Adaptive Signal Processing: Theory and Applications*, Springer-Verlag, New York, 1986.

B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

S. P. Applebaum and D. J. Chapman, "Adaptive arrays with main beam constraints," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, no. 5, pp. 650–662, September 1976.

R. B. Ash, *Basic Probability Theory*, John Wiley & Sons, New York, 1970.

R. B. Ash, *Real Analysis and Probability*, Academic Press, New York, 1972.

K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory*, Springer, New York, 2006.

C. Au and J. Tam, "Transforming variables using the Dirac generalized function," *The American Statistician*, vol. 53, no. 3, pp. 270–272, August 1999.

A. V. Balakrishnan, *Introduction to Random Processes in Engineering*, John Wiley & Sons, New York, 1995.

N. Balakrishnan, Ed., *Handbook of the Logistic Distribution*, Marcel Dekker, New York, 1992.

M. G. Bellanger, *Adaptive Digital Filters: Signal Processing and Communications*, 2nd edition, Marcel Dekker, New York, 2001.

S. Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind Deconvolution*, S. Haykin, Ed., Prentice-Hall, Englewood Cliffs, NJ, pp. 8–59, 1994.

A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Transactions on Communications*, vol. COM-32, no. 8, pp. 871–883, August 1984.

J. C. M. Bermudez and N. J. Bershad, "Transient and tracking performance analysis of the quantized LMS algorithm for time-varying system identification," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1990–1997, August 1996.

N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 4, pp. 793–806, August 1986.

D. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, Athena Scientific, Belmont, MA, 2002.

W. H. Beyer, *Standard Mathematical Tables*, 24th edition, CRC Press, Cleveland, OH, 1976.

P. Billingsley, *Probability and Measure*, 2nd edition, John Wiley & Sons, New York, 1986.

R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA, 1987.

G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting, and Control*, 4th edition, John Wiley & Sons, Hoboken, NJ, 2008.

S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

R. N. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, New York, 1978.

C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tip, "Acoustic echo control: An application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, July 1999.

Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 5, pp. 1081–1089, October 1986.

S. Brown and Z. Vranesic, *Fundamentals of Digital Logic with Verilog Design*, 2nd edition, McGraw-Hill, New York, 2008.

K. M. Buckley, "Spatial/spectral filtering with linearly constrained minimum variance beamformers," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 3, pp. 249–266, March 1987.

J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 907–939, September 1982.

J. A. Cadzow, *Foundations of Digital Signal Processing and Data Analysis*, Macmillan, New York, 1987.

R. E. Cagley, J. J. Shynk, and R. P. Gooch, "SINR enhancement in an IS-95 dowlink SIC receiver," *Wireless Personal Communications*, vol. 41, no. 4, pp. 563–580, June 2007.

M. Capiński and T. Zastawniak, *Probability Through Problems*, Springer, New York, 2001.

J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, August 1969.

G. F. Carrier and C. E. Pearson, *Partial Differential Equations: Theory and Techniques*, Academic Press, New York, 1976.

J. A. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least mean mixed-norm adaptive filtering," *Electronics Letters*, vol. 30, no. 19, pp. 1574–1575, September 1994.

S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 1, pp. 235–252, January 1998.

L. F. Chaparro, *Signals and Systems Using MATLAB®*, Academic Press, Burlington, MA, 2011.

C. Chatfield, *The Analysis of Time Series: An Introduction*, 2nd edition, Chapman and Hall, New York, 1980.

C. Chatterjee, Z. Kang, and V. P. Roychowdjury, "Algorithms for accelerated convergence of adaptive PCA," *IEEE Transactions on Neural Networks*, vol. 11, no. 2, pp. 338–355, March 2000.

R. S. Chhikara and J. L. Folks, *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*, Marcel Dekker, New York, 1989.

D. G. Childers, *Probability and Random Processes: Using MATLAB with Applications in Continuous and Discrete Time Systems*, Irwin/McGraw-Hill, Boston, 1997.

J. M. Cioffi, "Limited-precision effects in adaptive filtering," *IEEE Transactions on Circuits and Systems*, vol. CAS-34, no. 7, pp. 821–833, July 1987.

L. Cohen, "The history of noise," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 20–45, November 2005.

R. T. Compton, Jr., *Adaptive Antennas: Concepts and Performance*, John Wiley & Sons, Englewood Cliffs, NJ, 1988.

F. Conway and J. Siegelman, *Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics*, Basic Books, New York, 2005.

T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd edition, John Wiley & Sons, Hoboken, NJ, 2006.

H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ, 1946.

W. B. Davenport, Jr., and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.

W. B. Davenport, Jr., *Probability and Random Processes: An Introduction for Applied Scientists and Engineers*, McGraw-Hill, New York, 1970.

H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd edition, John Wiley & Sons, Hoboken, NJ, 2003.

R. de Neufville and J. H. Stafford, *Systems Analysis for Engineers and Managers*, McGraw-Hill, New York, 1971.

M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, 4th edition, Addison-Wesley, Reading, MA, 2011.

Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "Global convergence issues with linear blind adaptive equalizers," in *Blind Deconvolution*, S. Haykin, Ed., Prentice-Hall, Englewood Cliffs, NJ, pp. 60–120, 1994.

P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 3rd edition, Springer, New York, 2008.

P. M. Djurić, "Spatial spectrum estimation," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. J. R. Liu, Eds., John Wiley & Sons, Hoboken, NJ, 2009.

J. L. Doob, *Stochastic Processes*, John Wiley & Sons, New York, 1953.

J. L. Doob, *Measure Theory*, Springer-Verlag, New York, 1993.

R. C. Dorf, *Modern Control Systems*, 2nd edition, Addison-Wesley, Reading, MA, 1974.

E. R. Dougherty, *Probability and Statistics for the Engineering, Computing, and Physical Sciences*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, Cambridge, 2012.

H. Fan, "A structural view of asymptotic convergence speed of adaptive IIR filtering algorithms: Part I—Infinite precision implementation," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1493–1517, April 1993.

B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, John Wiley & Sons, New York, 1998.

G. Farmelo, *The Strangest Man: The Hidden Life of Paul Dirac, Mystic of the Atom*, Basic Books, New York, 2009.

W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. I, 3rd edition, John Wiley & Sons, New York, 1968.

W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, New York, 1971.

A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 1, pp. 222–230, February 1985.

T. L. Fine, *Probability and Probabilistic Reasoning for Electrical Engineering*, Prentice-Hall, Upper Saddle River, NJ, 2005.

T. E. Forster, *Set Theory with a Universal Set: Exploring an Untyped Universe*, 2nd edition, Clarendon Press, Oxford, 1995.

G. F. Franklin and J. D. Powell, *Digital Control of Dynamic Systems*, Addison-Wesley, Reading, MA, 1980.

B. Friedlander, "Lattice filters for adaptive processing," *Proceedings of the IEEE*, vol. 70, no. 8, pp. 829–867, August 1982.

O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, August 1972.

W. F. Gabriel, "Adaptive arrays: An introduction," *Proceedings of the IEEE*, vol. 64, no. 2, pp. 239–272, February 1976.

W. F. Gabriel, "Spectral analysis and adaptive array superresolution techniques," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 654–666, June 1980.

R. Gagliardi, *Introduction to Communications Engineering*, John Wiley & Sons, New York, 1978.

R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.

W. A. Gardner, *Introduction to Random Processes with Applications to Signals & Systems*, 2nd edition, McGraw-Hill, New York, 1990.

A. Gelb, Ed., *Applied Optimal Estimation*, MIT Press, Cambridge, 1974.

A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Transactions on Information Theory*, vol. IT-30, no. 2, pp. 191–199, March 1984.

G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., *Signal Processing Advances in Wireless and Mobile Communications, Volume 1: Trends in Channel Estimation and Equalization*, Prentice-Hall, Upper Saddle River, NJ, 2000.

J. D. Gibson, "Adaptive prediction in speech differential encoding systems," *Proceedings of the IEEE*, vol. 68, no. 4, pp. 488–525, April 1980.

J. D. Gibson, Ed., *The Communications Handbook*, CRC Press, Boca Raton, FL, 1997.

J. D. Gibson, T. Berger, T. Lookabaugh, D. Lindbergh, and R. L. Baker, *Digital Compression for Multimedia: Principles and Standards*, Morgan Kaufmann, San Francisco, 1998.

W. J. Gilbert, *Modern Algebra with Applications*, John Wiley & Sons, New York, 1976.

R. M. Glorioso, *Engineering Cybernetics*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

J. R. Glover, Jr., "Adaptive noise cancelling applied to sinusoidal interferences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-25, no. 6, pp. 484–491, December 1977.

D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, vol. COM-28, no. 11, pp. 1867–1875, November 1980.

M. A. Golberg, *An Introduction to Probability Theory with Statistical Applications*, Plenum Press, New York, 1984.

A. Goldsmith, *Wireless Communications*, Cambridge University Press, Cambridge, 2005.

G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1983.

R. P. Gooch and J. D. Lundell, "The CM array: An adaptive beamformer for constant modulus signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, pp. 2523–2526, April 1986.

S. E. Goodman and S. T. Hedetniemi, *Introduction to the Design and Analysis of Algorithms*, McGraw-Hill, New York, 1977.

G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction, and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, 1980.

R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag, New York, 1987.

R. M. Gray, *Entropy and Information*, Springer-Verlag, New York, 1990.

R. M. Gray and L. D. Davisson, *An Introduction to Statistical Signal Processing*, Cambridge University Press, Cambridge, 2004.

R. M. Gray and J. G. Goodman, *Fourier Transforms: An Introduction for Engineers*, Springer, New York, 1995.

M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice*, 3rd edition, John Wiley & Sons, Hoboken, NJ, 2008.

L. J. Griffiths, "A simple adaptive algorithm for real-time processing in antenna arrays," *Proceedings of the IEEE*, vol. 57, no. 10, pp. 1696–1704, October 1969.

L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. AP-30, no. 1, pp. 27–34, January 1982.

V. S. Groza and S. Shelley, *Precalculus Mathematics*, Holt, Rinehart, and Winston, New York, 1972.

J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge University Press, Cambridge, 2006.

E. J. Gumbel, "Bivariate exponential distributions," *Journal of the American Statistical Society*, vol. 55, no. 292, pp. 698–707, December 1960.

B. Hajek, *Notes for ECE 534: An Exploration of Random Processes for Engineers*, unpublished notes: http://www.ifp.illinois.edu/~hajek/Papers/randomprocJan09.pdf, 2009.

A. Hald, *Statistical Theory with Engineering Applications*, John Wiley & Sons, New York, 1952.

R. W. Hamming, *Digital Filters*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1983.

B. Hassibi, A. H. Sayed, and T. Kailath, "$\mathcal{H}^\infty$ optimality of the LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 267–280, February 1996.

D. Hatzinakos and C. L. Nikias, "Blind equalization using a tricepstrum-based algorithm," *IEEE Transactions on Communications*, vol. 39, no. 5, pp. 669–682, May 1991.

S. Haykin, *Adaptive Filter Theory*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1991.

S. Haykin *Advances in Spectrum Analysis and Array Processing*, vol. II, Prentice-Hall, Englewood Cliffs, NJ, 1991.

S. Haykin, Ed., *Blind Deconvolution*, Prentice-Hall, Englewood Cliffs, NJ, 1994.

S. Haykin, *Adaptive Filter Theory*, 4th edition, Prentice-Hall, Upper Saddle River, NJ, 2002.

S. Haykin, *Communication Systems*, fourth edition, John Wiley & Sons, New York, 2001.

S. Haykin and M. Moher, *Introduction to Analog and Digital Communications*, 2nd edition, John Wiley & Sons, Hoboken, NJ, 2007.

M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications*, Kluwer Academic Publishers, Boston, 1984.

C. W. Helstrom, *Probability and Stochastic Processes for Engineers*, 2nd edition, Macmillan, New York, 1991.

C. W. Helstrom, *Elements of Signal Detection and Estimation*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

F. S. Hillier and G. J. Lieberman, *Operations Research*, 2nd edition, Holden-Day, San Francisco, 1974.

P. G. Hoel, S. C. Port, and C. J. Stone, *Introduction to Stochastic Processes*, Houghton Mifflin, Boston, 1972.

M. L. Honig, S. L. Miller, M. J. Shensa, and L. B. Milstein, "Performance of adaptive linear interference suppression in the presence of dynamic fading," *IEEE Transactions on Communications*, vol. 49, no. 4, pp. 635–645, April 2001.

L. L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Transactions on Circuits and Systems*, vol. CAS-28, no. 6, pp. 562–576, June 1981.

Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Communication*, vol. 49, no. 7, pp. 588–601, July 2007.

S. Hwang and J. J. Shynk, "Blind GPS receiver with a modified despreader for interference suppression," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 2, pp. 503–513, April 2006.

O. C. Ibe, *Fundamentals of Applied Probability and Random Processes*, Elsevier Academic Press, London, 2005.

R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filtering," *IEEE Transactions on Communications*, vol. 42, no. 2/3/4, pp. 1017–1032, February/March/April 1994.

V. K. Ingle and J. G. Proakis, *Digital Signal Processing Using MATLAB®*, 3rd edition, Cengage Learning, Stamford, CT, 2011.

Y. Isukapalli and B. D. Rao, "An analytically tractable approximation for the Gaussian *Q*-function," *IEEE Communications Letters*, vol. 12, no. 9, pp. 669–671, September 2008.

N. K. Jablon, "Steady state analysis of the generalized sidelobe canceller by adaptive noise cancelling techniques," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 330–337, March 1986.

W. C. Jakes, Ed., *Microwave Mobile Communications*, Wiley-IEEE Press, New York, 1994.

N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

C. R. Johnson, Jr., "Adaptive IIR filtering: Current results and open issues," *IEEE Transactions on Information Theory*, vol. IT-30, no. 2, pp. 237–250, March 1984.

C. R. Johnson, Jr., *Lectures on Adaptive Parameter Estimation*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927–1950, October 1998.

N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1, 2nd edition, John Wiley & Sons, New York, 1994.

N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 2, 2nd edition, John Wiley & Sons, New York, 1995.

N. L. Johnson, S. Kotz, and N. Balakrishnan, *Discrete Multivariate Distribution*, John Wiley & Sons, New York, 1997.

N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distributions*, 3rd edition, John Wiley & Sons, Hoboken, NJ, 2005.

I. T. Jolliffe, "A note on the use of principal components in regression," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 31, no. 3, pp. 300–303, 1982.

I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

T. Kailath, "A view of three decades of linear filtering theory," *IEEE Transactions on Information Theory*, vol. IT-20, no. 2, pp. 146–181, March 1974.

T. Kailath, *Lectures on Linear Least-Squares Estimation*, Springer-Verlag, New York, 1976.

T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, Upper Saddle River, NJ, 2000.

R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, series D, pp. 34–45, March 1960.

R. E. Kalman and R. S. Bucy, "New results in linear filtering an prediction theory," *Transactions of the ASME–Journal of Basic Engineering*, vol. 83, series D, pp. 95–108, March 1961.

S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd edition, Academic Press, New York, 1975.

S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, November 1981.

S. M. Kay, *Modern Spectral Estimation: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1998.

A. V. Keerthi, A. Mathur, and J. J. Shynk, "Misadjustment and tracking analysis of the constant modulus array," *IEEE Transactions on Signal Processing* vol. 46, no. 1, pp. 51–58, January 1998.

J. F. C. Kingman and S. J. Taylor, *Introduction to Measure and Probability*, Cambridge University Press, Cambridge, 1966.

B. Kolman, *Elementary Linear Algebra*, 2nd edition, Macmillan, New York, 1977.

A. N. Kolmogorov, *Foundations of the Theory of Probability*, 2nd edition, Chelsea, New York, 1960.

B. Kosko, Ed., *Neural Networks for Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1992.

S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions: Models and Applications*, vol. 1, John Wiley & Sons, New York, 2000.

S. Kotz, T. J. Kozubowski, and K. Podgórski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Birkhäuser, Boston, 2001.

E. Kreyszig, *Advanced Engineering Mathematics*, 4th edition, John Wiley & Sons, New York, 1979.

H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

K. Krishnamoorthy, *Handbook of Statistical Distributions with Applications*, Chapman & Hall/CRC Press, Boca Raton, FL, 2006.

V. Krishnan, *Probability and Random Processes*, John Wiley & Sons, Hoboken, NJ, 2006.

M. Krstić, I. Kanellakopoulos, and P. Kokotović, *Nonlinear and Adaptive Control Design*, John Wiley & Sons, New York, 1995.

A. Kuh, "Comparison of tracking algorithms for single layer threshold networks in the presence of random drift," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 640–649, March 1997.

S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, March 1951.

R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no. 1, pp. 134–139, January 1983.

M. Kunt, *Digital Signal Processing*, Artech House, Norwood, MA, 1987.

K.-C. Lai and J. J. Shynk, "Performance evaluation of a generalized linear SIC for DS/CDMA signals," *IEEE Transactions on Signal Processing*, vol. 51, no. 6, pp. 1604–1614, June 2003.

I. D. Landau, *Adaptive Control: The Model Reference Approach*, Marcel Dekker, New York, 1979.

H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences: Random Variables and Stochastic Processes*, vol. I, John Wiley & Sons, New York, 1979.

H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences: Random Noise, Signals and Dynamic Systems*, vol. II, John Wiley & Sons, New York, 1979.

B. P. Lathi, *Signals, Systems, and Communication*, John Wiley & Sons, New York, 1965.

B. P. Lathi, *Signal Processing & Linear Systems*, Oxford University Press, New York, 2000.

B. P. Lathi and Z. Ding, *Modern Digital and Analog Communication Systems*, 4th edition, Oxford University Press, New York, 2009.

A. J. Laub, *Matrix Analysis for Scientists and Engineers*, Society for Industrial and Applied Mathematics, Philadelphia, 2005.

W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill, New York, 1989.

E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd edition, Kluwer Academic Publishers, Boston, 1994.

E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley & Sons, New York, 1959.

E. L. Lehmann, *Theory of Point Estimation*, John Wiley & Sons, New York, 1983.

S. J. Leon, *Linear Algebra with Applications*, 3rd edition, Macmillan, New York, 1990.

F. C. Leone, L. S. Nelson, and R. B. Nottingham, "The folded normal distribution," *Technometrics*, vol. 3, no. 4, pp. 543–550, November 1961.

A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineers*, Prentice-Hall, Upper Saddle River, NJ, 2008.

H. Lev-Ari, T. Kailath, and J. M. Cioffi, "Least-squares adaptive lattice and transversal filters: A unified geometric theory," *IEEE Transactions on Information Theory*, vol. IT-30, no. 2, pp. 222–236, March 1984.

B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, Springer, New York, 2008.

Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2120–2129, September 1995.

M. Liebling, T. Blu, and M. Unser, "Fresnelets: New multiresolution wavelet bases for digital holography," *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 29–43, January 2003.

E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: Keys and clues," *BioScience*, vol. 51, no. 5, pp. 341–352, May 2001.

R. E. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol. 4, no. 2, pp. 4–22, April 1987.

L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.

R. Lucky, "Automatic equalization for digital communications," *Bell System Technical Journal*, vol. 44, no. 4, pp. 547–588, April 1965.

L. C. Ludeman, *Random Processes: Filtering, Estimation, and Detection*, John Wiley & Sons, Hoboken, NJ, 2003.

D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd edition, Addison-Wesley, Reading, MA, 1984.

J. Lundell and B. Widrow, "Applications of the constant modulus adaptive beamformer to constant and non-constant modulus signals," in *Proceedings of the Twenty-First Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, pp. 432–436, November 1987.

D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, 2003.

U. Madhow, *Fundamentals of Digital Communication*, Cambridge University Press, Cambridge, 2008.

V. Madisetti, Ed., *The Digital Signal Processing Handbook*, 2nd edition, CRC Press, Boca Raton, FL, 2009.

J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, April 1975.

B. S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG-7: Mulimedia Content Description Interface*, John Wiley & Sons, New York, 2002.

D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*, McGraw-Hill, New York, 2000.

K. V. Mardia, *Families of Bivariate Distributions*, Griffin, London, 1970.

J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.

G. Marsaglia, "Ratios of normal variables," *Journal of Statistical Software*, vol. 16, no. 4, pp. 1–10, May 2006.

V. J. Mathews, "Performance analysis of adaptive filters equipped with the dual sign algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 1, pp. 85–91, January 1991.

J. H. McClellan, R. W. Schafer, and M. A. Yoder, *DSP First: A Multimedia Approach*, Prentice-Hall, Upper Saddle River, NJ, 1998.

R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication*, Addison-Wesley, Reading, MA, 1977.

E. L. McMahon, "An extension of Price's theorem," *IEEE Transactions on Information Theory*, vol. 10, no. 2, p. 168, April 1964.

E. L. Melnick and A. Tenenbein, "The misspecifications of the normal distribution," *The American Statistician*, vol. 36, no. 4, pp. 372–373, November 1982.

Members of the Technical Staff, Bell Laboratories, *Transmission Systems for Communications*, 4th edition, Bell Telephone Laboratories/Western Electric, Winston-Salem, NC, 1970.

J. M. Mendel, *Discrete Techniques of Parameter Estimation: The Equation Error Formulation*, Marcel Dekker, New York, 1973.

J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and systems theory: Theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, March 1991.

J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

D. Middleton, *An Introduction to Statistical Communication Theory*, Wiley-IEEE Press, New York, 1996.

I. Miller and J. E. Freund, *Probability and Statistics for Engineers*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1977.

T. M. Mills, *Problems in Probability*, World Scientific Publishing Company, River Edge, NJ, 2001.

M. L. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*, expanded edition, MIT Press, Cambridge, MA, 1988.

S. K. Mitra and J. F. Kaiser, Eds., *Handbook for Digital Signal Processing*, John Wiley & Sons, New York, 1993.

S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 4th edition, McGraw-Hill, New York, 2010.

N. Mohanty, *Random Signals, Estimation, and Identification: Analysis and Applications*, Van Nostrand Reinhold, New York, 1986.

R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, John Wiley & Sons, New York, 1980.

A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, 3rd edition, McGraw-Hill, New York, 1974.

T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, November 1996.

T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice-Hall, Upper Saddle, NJ, 2000.

F. Mosteller and J. W. Tukey, *Data Analysis and Regression: A Second Course in Statistics*, Addison-Wesley, Reading, MA, 1977.

B. Mulgrew and C. F. N. Cowan, *Adaptive Filters and Equalisers*, Kluwer Academic Publishers, Boston, 1988.

H. T. Nagle, Jr., B. D. Carroll, and J. D. Irwin, *An Introduction to Computer Logic*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

M. Nayeri, H. Fan, and W. K. Jenkins, "Some Characteristics of error surfaces for insufficient order adaptive IIR filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1222–1227, July 1990.

C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice-Hall, Upper Saddle River, NJ, 1993.

J. W. Nilsson and S. A Riedel, *Electric Circuits*, 7th edition, Pearson Prentice-Hall, Upper Saddle River, NJ, 2005.

J. J. O'Connor and E. F. Robertson, *The MacTutor History of Mathematics Archive*, http://www-history.mcs.st-andrews.ac.uk/index.html, 2011.

T. Ogunfunmi and M. Narasimha, *Principles of Speech Coding*, CRC Press, Boca Raton, FL, 2010.

P. Olofsson, *Probability, Statistics, and Stochastic Processes*, John Wiley & Sons, Hoboken, NJ, 2005.

A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ 1975.

A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1996.

A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd edition, Prentice-Hall, Upper Saddle River, NJ, 2009.

N. Pal, C. Jin, and W. K. Lim, *Handbook of Exponential and Related Distributions for Engineers and Scientists*, Chapman & Hall/CRC Press, Boca Raton, FL, 2006.

A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

A. Papoulis, *Probability & Statistics*, Prentice-Hall, Englewood, Cliffs, NJ, 1990.

E. Parzen, *Modern Probability Theory and Its Applications*, John Wiley & Sons, New York, 1960.

E. Parzen, *Stochastic Processes*, Holden-Day, San Francisco, 1962.

R. F. Pawula, "A modified version of Price's theorem," *IEEE Transactions on Information Theory*, vol. IT-13, no. 2, pp. 285–288, April 1967.

P. Z. Peebles, Jr., *Probability, Random Variables, and Signal Principles*, 4th edition, McGraw-Hill, New York, 2001.

M. R. Petraglia, R. G. Alves, and P. S. R. Diniz, "New structures for adaptive filtering in subbands with critical sampling," *IEEE Transactions on Signal Processing*, vol. 48, no 12, pp. 3316–3327, December 2000.

G. Picchi and G. Prati, "Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm," *IEEE Transactions on Communications*, vol. COM-35, no. 9, pp. 877–887, September 1987.

S. U. Pillai, T. Suel, and S. Cha, "The Perron-Frobenius theorem," *IEEE Signal Processing Magazine*, vol. 22, no. 2, pp. 62–75, March 2005.

K. C. Pohlmann, *Principles of Digital Audio*, 4th edition, McGraw-Hill, New York, 2000.

C. Polprasert and J. A. Ritcey, "A Nakagami fading phase difference distribution and its impact on BER performance," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2805–2813, July 2008.

H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd edition, Springer, New York, 1994.

H. V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, Cambridge, 2008.

B. Porat, *Digital Processing of Random Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1994.

R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Transactions on Information Theory*, vol. 4, no. 2, pp. 69–72, June 1958.

J. G. Proakis, *Digital Communications*, fourth edition, McGraw-Hill, New York, 2001.

J. G. Proakis and M. Salehi, *Communication Systems Engineering*, Prentice-Hall, Upper Saddle River, NJ, 2002.

P. E. Protter, *Stochastic Integration and Differential Equations*, 2nd edition, Springer, New York, 2004.

M. B. Pursley, *Random Processes in Linear Systems*, Prentice-Hall, Upper Saddle River, NJ, 2002.

M. B. Pursley, *Introduction to Digital Communications*, Pearson Prentice-Hall, Upper Saddle River, NJ, 2005.

S. U. H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349–1387, September 1985.

L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.

H. Raiffa, *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley, Reading, MA, 1970.

T. S. Rappaport, *Wireless Communications*, Prentice-Hall, Upper Saddle River, NJ, 1996.

I. S. Reed, "On a moment theorem for complex Gaussian processes" *IRE Transactions on Information Theory*, vol. 8, no. 3, pp. 194–195, April 1962.

P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, New York, 1995.

S. O. Rice, "Mathematical analysis of random noise," *Bell System Technical Journal*, vol. 24, no. 1, pp. 46–156, January 1945.

L. F. Richardson, *Measure and Integration: A Concise Introduction to Real Analysis*, John Wiley & Sons, Hoboken, NJ, 2009.

E. A. Robinson, *Probability Theory and Applications*, International Human Development Corporation, Boston, 1985.

K. Rose and S. L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 965–976, July 2001.

K. A. Ross, *Elementary Analysis: The Theory of Calculus*, Springer-Verlag, New York, 1980.

S. Ross, *A First Course in Probability*, MacMillan, New York, 1976.

S. M. Ross, *Stochastic Processes*, John Wiley & Sons, New York, 1983.

S. A. Ross, *Introduction to Ordinary Differential Equations*, 2nd edition, John Wiley & Sons, New York, 1977.

D. W. Rousu, "Weibull skewness and kurtosis as a function of the shape parameter," *Technometrics*, vol. 15, no. 4, pp. 927–930, November 1973.

S. Roy, "Subspace blind adaptive detection for multiuser CDMA," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 169–175, January 2000.

M. Rupp and A. H. Sayed, "On the convergence of blind adaptive equalizers for constant modulus signals," *IEEE Transactions on Communications*, vol. 48, no. 5, pp. 795–803, May 2000.

D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, MIT Press, Cambridge, MA, 1986.

J. Salz, Optimum mean-square decision feedback equalization," *Bell System Technical Journal*, vol. 52, no. 8, pp. 1341–1373, October 1973.

Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Transactions on Communications*, vol. COM-23, no. 6, pp. 679–682, June 1975.

A. H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, Hoboken, NJ, 2003.

A. H. Sayed, *Adaptive Filters*, John Wiley & Sons, Hoboken, NJ, 2008.

L. L. Scharf, *Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

R. J. Schilling and H. Lee, *Engineering Analysis: A Vector Space Approach*, John Wiley & Sons, New York, 1988.

R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 276–281, March 1986.

T. A. Schonhoff and A. A. Giordano, *Detection and Estimation Theory and Its Applications*, Prentice-Hall, Upper Saddle River, NJ, 2006.

M. Schwartz, *Information, Transmission, Modulation, and Noise*, 3rd edition, McGraw-Hill, New York, 1980.

V. Seshadri, *The Inverse Gaussian Distribution: Statistical Theory and Applications*, Springer-Verlag, New York, 1999.

W. A. Sethares, "Adaptive algorithms with nonlinear data and error functions," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2199–2206, September 1992.

O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 312–321, March 1990.

K. S. Shanmugan and A. M. Breipohl, *Random Signals: Detection, Estimation, and Data Analysis*, John Wiley & Sons, New York, 1988.

C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, July 1948.

C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, October 1948.

C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1963.

A. N. Shiryayev, *Probability*, Springer-Verlag, New York, 1984.

J. J. Shynk, "Adaptive IIR Filtering," *IEEE ASSP Magazine*, vol. 6, no. 2, pp. 4–21, April 1989.

J. J. Shynk and C. K. Chan, "Performance surfaces of the constant modulus algorithm based on a conditional Gaussian model," *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1965–1969, May 1993.

J. J. Shynk and R. P. Gooch, "The constant modulus array for cochannel signal copy and direction finding" *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 652–660, March 1996.

M. K. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers, Scientists, and Mathematicians*, Springer, New York, 2002.

D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2811–2825, September 1993.

I. S. Sokolnikoff and R. M. Redheffer, *Mathematics of Physics and Modern Engineering*, McGraw-Hill, New York, 1966.

V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

H. W. Sorenson, *Parameter Estimation: Principles and Problems*, Marcel Dekker, New York, 1980.

M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to Statistical Signal Processing with Applications*, Prentice-Hall, Upper Saddle River, NJ, 1996.

H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, 3rd edition, Prentice-Hall, Upper Saddle River, NJ, 2002.

S. D. Stearns and D. R. Hush, *Digital Signal Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Upper Saddle River, NJ, 1997.

P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

G. Strang, *Linear Algebra and Its Applications*, Academic Press, New York, 1976.

P. Strobach, *Linear Prediction Theory: A Mathematical Basis for Adaptive Systems*, Springer-Verlag, New York, 1990.

B. L. Sturm and J. J. Shynk, "Sparse approximation and the pursuit of meaningful signal models with interference adaptation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 461–472, March 2010.

A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors, part I: The MUSIC algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1758–1774, July 1992.

M. E. Tarter and V. A. Clark, "Properties of the median and other order statistics of logistic variates," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1779–1786, December 1965.

H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, Academic Press, New York, 1984.

C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1992.

C. W. Therrien and M. Tummala, *Probability and Random Processes for Electrical and Computer Engineers*, 2nd edition, CRC Press, Boca Raton, FL, 2012.

R. E. Thomas, A. J. Rosa, and G. J. Toussaint, *The Analysis and Design of Linear Circuits*, 6th edition, John Wiley & Sons, Hoboken, NJ, 2009.

G. B. Thomas, Jr., *Calculus and Analytic Geometry, Part One: Functions of One Variable and Analytic Geometry*, 4th edition, Addison-Wesley, Reading, MA, 1968.

G. B. Thomas, Jr., *Calculus and Analytic Geometry, Part Two: Vectors and Functions of Several Variables*, 4th edition, Addison-Wesley, Reading, MA, 1968.

L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time-domain approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340–349, March 1994.

J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 2, pp. 459–472, April 1983.

J. R. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters*, Prentice-Hall, Upper Saddle River, NJ, 2001.

S. A. Tretter, *Introduction to Discrete-Time Signal Processing*, John Wiley & Sons, New York, 1976.

J. K. Tugnait, "On fractionally spaced blind adaptive equalization under symbol timing offsets using Godard and related equalizers," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1817–1821, July 1996.

P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, John Wiley & Sons, New York, 1968.

H. L. Van Trees, *Optimum Array Processing, Part IV of Detection Estimation, and Modulation Theory*, John Wiley & Sons, New York, 2002.

B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, pp. 4–24, April 1988.

V. Venkataraman and J. J. Shynk, "A multistage hybrid constant modulus array with constrained adaptation for correlated sources," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2509–2519, June 2007.

S. Verdú, *Multiuser Detection*, Cambridge University Press, Cambridge, 1998.

A. C. G. Verdugo Lazo and P. N. Rathie, "On the entropy of continuous probability distributions," *IEEE Transactions on Information Theory*, vol. IT-24, no. 1, pp. 120–122, January 1978.

Y. Viniotis, *Probability and Random Processes for Electrical Engineers*, McGraw-Hill, New York, 1998.

E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Transactions on Information Theory*, vol. IT-30, no. 2, pp. 275–283, March 1984.

E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics*, 2nd edition, Chapman & Hall/CRC Press, Boca Raton, FL, 2003.

B. Widrow, K. M. Duvall, R. P. Gooch, and W. C. Newman, "Signal cancellation phenomena in adaptive antennas: Causes and cures," *IEEE Transactions on Antennas and Propagation*, vol. AP-30, no. 3, pp. 469–478, May 1982.

B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ 1985.

B. Widrow and E. Walach, *Adaptive Inverse Control*, Prentice-Hall, Upper Saddle River, NJ, 1996.

B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: Perceptron, madaline, and backpropagation," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415–1442, September 1990.

B. Widrow and I. Kollár, *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*, Cambridge University Press, Cambridge, 2008.

N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, MIT Press, Cambridge, MA, 1949.

N. Wiener, *Time Series*, MIT Press, Cambridge, MA, 1964.

D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, 1991.

R. H. Williams, *Electrical Engineering Probability*, West Publishing, St. Paul, MN, 1991.

G. A. Williamson, "Implementations of adaptive IIR filters with lowest complexity," *IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing*, vol. 44, no. 8, pp. 673–678, August 1997.

E. Wong, *Stochastic Processes in Information and Dynamical Systems*, Robert E. Krieger Publishing Company, Huntington, NY, 1971.

E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, Springer-Verlag, New York, 1985.

J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, New York, 1965.

Z. Xu, P. Liu, and X. Wang, "Blind multiuser detection: From MOE to subspace methods," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 510–524, February 2004.

R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*, John Wiley & Sons, New York, 2005.

M.-H. Yeon and J. J. Shynk, "A stochastic analysis of an iterative semi-blind beamformer for TDMA systems," *Signal Processing*, vol. 89, no. 7, pp. 1347–1358, July 2009.

J. R. Zeidler, "Performance analysis of LMS adaptive prediction filters," *Proceedings of the IEEE*, vol. 78, no. 12, pp. 1781–1806, December 1990.

D. Zwillinger and S. Kokoska, *CRC Standard Probability and Statistics Tables and Formulae*, Chapman & Hall/CRC Press, Boca Raton, FL, 2000.

{全文地址} http://202.206.108.43:8099/13/disksnu/snu23/03/

【作　者】（美）约翰 J·申克著；谢晓霞，安成锦，许可译
【形态项】531
【出版项】北京：机械工业出版社，2016.04
【ISBN号】978-7-111-51965-2
【原书定价】129.00
【中图法分类号】TN911.7